Karlsruher Institut für Technologie

Signalverarbeitung in der Nachrichtentechnik – Schätztheorie

Holger Jäkel

Communications Engineering Lab (CEL)

Übersicht

3. Schätztheorie
   - Einführung und Grundlagen
   - Least-Squares-Schätzung
   - Eigenschaften von Schätzern
   - MMSE-Schätzer
   - Verteilungsbasierte Schätzer
   - Lernzielkontrolle
   - Literatur

Literatur

Folgende Betrachtungen und Herleitungen basieren größtenteils auf:

H. Kronmüller, *Digitale Signalverarbeitung*, Springer, 1991

Übersicht

Einführung – Begrifflichkeiten

- *Schätzung*
  - Bestimmen eines unbekannten Parametervektors $\boldsymbol{b} \in \mathbb{C}^K$
  - Basierend auf Beobachtung $\boldsymbol{y} \in \mathbb{C}^N$ ergibt sich die Abfolge[1]

$$\boldsymbol{b} \rightarrow \boldsymbol{y} \rightarrow \hat{\boldsymbol{b}} = S\{\boldsymbol{y}\}$$

  - **Anwendung:**
    - Bestimmung von Trägerphase und -frequenz; Kanalschätzung
    - Synchronisation

- *Detektion*
  - Vorher bekannt: $\boldsymbol{b} \in B \subseteq \mathbb{C}^K, |B| < \infty$
  - **Anwendung:**
    - Demodulation
    - Decodierung

[1] Die Schreibweise der Schätzer als $\hat{\boldsymbol{b}}$ oder $S\{\boldsymbol{y}\}$ wird angepasst gewählt, je nachdem, ob die Abhängigkeit von der Beobachtung oder der zu schätzende Wert hervorgehoben werden soll.

Lineares Signalmodell

- Lineares Signalmodell beschreibt einen linearen Zusammenhang[2] zwischen Beobachtung $\boldsymbol{y}$ und gesuchtem Parametervektor $\boldsymbol{b}$:

$$\boldsymbol{y} = \sum_{i=1}^{K} \boldsymbol{x}_k b_k + \boldsymbol{n} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{n}$$

mit

$$\begin{aligned} \boldsymbol{X} &= [\boldsymbol{x}_1 \ldots \boldsymbol{x}_K] \in \mathbb{C}^{N\times K} \\ \boldsymbol{b} &= (b_1, \ldots, b_K)^{\mathrm{T}} \in \mathbb{C}^K \\ \boldsymbol{y} &= (y_1, \ldots, y_N)^{\mathrm{T}} \in \mathbb{C}^N \\ \boldsymbol{n} &\in \mathbb{C}^N \sim f_{\boldsymbol{N}}(\boldsymbol{n}) \end{aligned}$$

[2] **Übung:** Interpretieren Sie diesen Ansatz im Hinblick auf Vektorraumtheorie.

Lineare Schätzung

- *Lineare Schätzung* schätzt den Parametervektor $\boldsymbol{b}$ mittels linearer Operation aus der Beobachtung $\boldsymbol{y}$, d.h. mittels einer Matrix $\boldsymbol{G} \in \mathbb{C}^{K \times N}$ gilt für das lineare Signalmodell:

$$\hat{\boldsymbol{b}} = \boldsymbol{G}\boldsymbol{y} = \boldsymbol{G}\boldsymbol{X}\boldsymbol{b} + \boldsymbol{G}\boldsymbol{n}$$

- **Bemerkung:** Schreibt man[3] mit $\boldsymbol{y}[n] := (y[n], \ldots, y[n-N+1])^{\mathrm{T}}$ die Schätzgleichung als

$$\hat{\boldsymbol{b}} = \boldsymbol{G}\boldsymbol{y}[n] = [\boldsymbol{g}_1, \ldots, \boldsymbol{g}_N]^{\mathrm{T}} \boldsymbol{y}[n] \quad \Rightarrow \quad \hat{b}_i = \boldsymbol{g}_i^{\mathrm{T}} \boldsymbol{y} = \sum_{m=0}^{N-1} (\boldsymbol{g}_i)_{m+1}\, y[n-m],$$

so fällt auf, dass es sich um eine Filterung mit einem kausalen, linearen Filter mit der Impulsantwort $\boldsymbol{g}_i$ handelt.

[3] Für diese Illustration werden die Indizes verschoben und eine Zeitrichtung „eingebaut“.

Gaußapproximation/Orthogonalitätsprinzip

Theorem (Orthogonalitätsprinzip)

Ist $B = \{x_1, \ldots, x_n\}$ eine ONB eines Teilraums $\emptyset \neq S \subseteq V$ in einem Innenproduktraum V, so ist für $y \in V$ die Gaußsche Approximationsaufgabe

$$\min_{\tilde{y} \in S} \|y - \tilde{y}\|$$

eindeutig durch $\alpha_i = \langle y, x_i \rangle$ und damit

$$y_a := \arg\min_{\tilde{y} \in S} \|y - \tilde{y}\| = \sum_{i=1}^{n} \langle y, x_i \rangle x_i$$

lösbar. Der Fehlervektor steht bei Erreichen des Minimums senkrecht auf S,

$$(y - y_a) \perp S.$$

Gaußapproximation/Orthogonalitätsprinzip

- **Bemerkung:**
  - Das Orthogonalitätsprinzip stellt einen möglichen Ausgangspunkt zur Bestimmung optimaler Schätzer dar.
  - Alternativ sind anderen Ansätze wie z.B.
    - Ableiten und Nullsetzen
    - Quadratische Ergänzung

  denkbar. Welche „Variante“ einfacher/effizienter ist, kann von Fall zu Fall variieren.

Übersicht

3 Schätztheorie

- Einführung und Grundlagen
- Least-Squares-Schätzung
- Eigenschaften von Schätzern
- MMSE-Schätzer
- Verteilungsbasierte Schätzer
- Lernzielkontrolle
- Literatur

Least-Squares-Schätzung I

- Die Schätzung des gesuchten Parameters ergibt einen Vektor $\hat{\boldsymbol{b}}$
- Aufgrund des Signalmodells liegt der „zurückgeschätzte“ Wert ohne Rauschen in dem von den Spalten von $\boldsymbol{X}$ aufgespannten Unterraum:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{b}}.$$

- Nach Gaußapproximation gilt für die bestmögliche Approximation notwendigerweise $(\boldsymbol{y} - \hat{\boldsymbol{y}}) \perp \boldsymbol{x}_k$ und somit folgt:[4]

$$\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\hat{\boldsymbol{b}} = \boldsymbol{X}^{\mathrm{H}}\boldsymbol{y}.$$

  Gleichungslösung liefert:

$$\hat{\boldsymbol{b}}_{\mathrm{LS}} = (\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{y}$$

[4] Dies sind die so genannten Normalengleichungen.

Least-Squares-Schätzung II

- Ziel ist die Minimierung der Fehlerquadrate:

$$\|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2 \stackrel{!}{=} \min$$

Einsetzen liefert:

$$\|\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{b}}\|^2 \stackrel{!}{=} \min$$

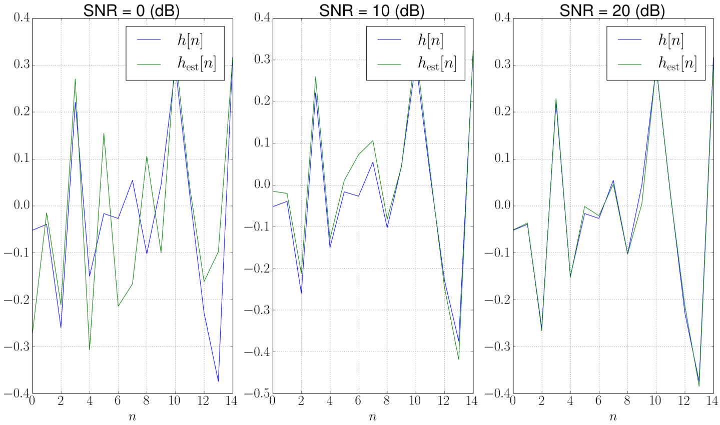
Die Lösung ergibt sich durch die *Pseudoinverse* einer Matrix und ist[5]:

$$\hat{\boldsymbol{b}}_{\mathrm{LS}} = \boldsymbol{X}^{\dagger}\boldsymbol{y} = (\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{y}$$

[5] ..., falls $\mathrm{rk}(\boldsymbol{X}) = K = \min\{N, K\}$, d. h. $\boldsymbol{X}$ vollen Spaltenrang hat und es weniger Spalten als Zeilen sind. Voller Spaltenrang entspricht der linearen Unabhängigkeit der Spalten. Sind die Spalten aber linear unabhängig, so ist notwendigerweise $K \leq N$. (**Übung:** Wieso?)

Least-Squares-Schätzung II

- **Beispiel:** LS-Schätzer für eine Impulsantwort der Länge 15; Eingangssignal ist eine BPSK-modulierte PN-Sequenz der Länge 31



Übersicht

3 Schätztheorie

- Einführung und Grundlagen
- Least-Squares-Schätzung
- Eigenschaften von Schätzern
- MMSE-Schätzer
- Verteilungsbasierte Schätzer
- Lernzielkontrolle
- Literatur

Eigenschaften von Schätzern

Definitionen

- Ein Schätzer $\hat{\boldsymbol{b}}$ heißt *erwartungstreu*, falls $E(\hat{\boldsymbol{b}}) = \boldsymbol{b}$.
- Falls der Schätzer $\hat{\boldsymbol{b}}$ bei beliebig langer Beobachtung beliebig genau schätzt, heißt er *konsistent*:

$$\boldsymbol{y} \in \mathbb{C}^N \Rightarrow \lim_{N\to\infty} P(|\hat{\boldsymbol{b}} - \boldsymbol{b}| > \epsilon) = 0$$

- Ein erwartungstreuer Schätzer $\hat{\boldsymbol{b}}_e$ heißt *effizient*, falls für jeden anderen Schätzer $\hat{\boldsymbol{b}}$ gilt:[6]

$$\boldsymbol{C}_{\hat{\boldsymbol{b}}_e\hat{\boldsymbol{b}}_e} \leq \boldsymbol{C}_{\hat{\boldsymbol{b}}\hat{\boldsymbol{b}}}$$

- **Beispiel:** Ein linearer Schätzer ist bei mittelwertfreier Störung erwartungstreu, falls $\boldsymbol{G}\boldsymbol{X} = \boldsymbol{I}$ ist.

[6] Man definiert für Matrizen: $\boldsymbol{A} \leq \boldsymbol{B}$, falls $\boldsymbol{B} - \boldsymbol{A}$ positiv semidefinit ist. Dazu ist äquivalent, dass $\boldsymbol{x}^{\mathrm{H}}(\boldsymbol{B} - \boldsymbol{A})\boldsymbol{x} \geq 0,\ \boldsymbol{x} \in \mathbb{C}^n$ gilt bzw. dass alle Eigenwerte von $\boldsymbol{B} - \boldsymbol{A}$ nicht-negativ sind.

Eigenschaften von Schätzern

- **Bemerkung:**
  - Die Forderung der Erwartungstreue in der Definition der Effizienz ist nicht notwendig, man kann Effizienz auch für Schätzer mit $E(\hat{\boldsymbol{b}}) \neq \boldsymbol{b}$ definieren. Dann ist die Varianz des Schätzer aber kein vernünftiges Gütemaß.[7]
    Für den einer Kovarianzmatrix nachempfundenen Ausdruck folgt:[8]

$$E((\hat{\boldsymbol{b}} - \boldsymbol{b})(\hat{\boldsymbol{b}} - \boldsymbol{b})^{\mathrm{H}}) = \boldsymbol{C}_{\hat{\boldsymbol{b}}\hat{\boldsymbol{b}}} + (\boldsymbol{b} - E(\hat{\boldsymbol{b}}))(\boldsymbol{b} - E(\hat{\boldsymbol{b}}))^{\mathrm{H}}.$$

    Durch Bildung der Spur folgt für die *mittlere quadratische Abweichung, mean square error, MSE*:

$$\begin{aligned} \mathrm{MSE} &:= E(\|\hat{\boldsymbol{b}} - \boldsymbol{b}\|^2) \\ &= E(\|\hat{\boldsymbol{b}} - E(\hat{\boldsymbol{b}})\|^2) + \|\boldsymbol{b} - E(\hat{\boldsymbol{b}})\|^2 \end{aligned}$$

    anstelle der Varianz des Schätzers.[9]

---

[7] **Übung:** Finden Sie ein Gegenbeispiel, das diese Aussage unterstreicht.

[8] **Übung**: Nachweis

[9] In beiden Gleichungen unterscheiden sich die Ausdrücke von der „normalen (Ko-)Varianz“ durch den systematischen Fehler, der durch die fehlende Erwartungstreue entsteht.

Eigenschaften von Schätzern

- **Beispiel:** (Mehrfache Messung eines Wertes)
  - Modell $\boldsymbol{y} = (1, \dots, 1)^{\mathrm{T}} b + \boldsymbol{n}, b \in \mathbb{R}, \boldsymbol{y}, \boldsymbol{n} \in \mathbb{R}^N$
  - Der lineare Schätzer mit $\boldsymbol{G} = (1/N, \dots, 1/N) \in \mathbb{R}^{1 \times N}$ resultiert in

$$\hat{b} = \frac{1}{N} \sum_{n=1}^{N} y_i = b + \frac{1}{N} \sum_{n=1}^{N} n_i,$$

    ist damit erwartungstreu und konsistent (**Übung**; Hinweis: Tschebyscheff-Ungleichung).
  - Ist der Schätzer effizient? Wie könnte man das überprüfen?

Cramer-Rao-Schranke

Theorem (Cramer-Rao)[10]

Ist der Schätzer $\hat{\boldsymbol{b}}$ erwartungstreu, so gilt[11] für die Kovarianzmatrix des Schätzfehlers $\tilde{\boldsymbol{b}} := \boldsymbol{b} - \hat{\boldsymbol{b}}$:

$$\boldsymbol{C}_{\tilde{\boldsymbol{b}}\tilde{\boldsymbol{b}}} \geq \boldsymbol{J}^{-1}$$

mit der *Fisher'schen Informationsmatrix*

$$\begin{aligned}\boldsymbol{J} &:= E\left(\left[\frac{\partial \ln(f(\boldsymbol{y}|\boldsymbol{b}))}{\partial \boldsymbol{b}}\right]\left[\frac{\partial \ln(f(\boldsymbol{y}|\boldsymbol{b}))}{\partial \boldsymbol{b}}\right]^{\mathrm{T}}\right) \\ &= -E\left(\frac{\partial^2 \ln(f(\boldsymbol{y}|\boldsymbol{b}))}{\partial \boldsymbol{b}^2}\right).\end{aligned}$$

[10] Zur Vereinfachung für reellwertige Größen formuliert und für den eindimensionalen Fall nachgewiesen.

[11] ...bei Existenz der Größen

Cramer-Rao-Schranke

- Ist insbesondere $b \in \mathbb{R}$, so lautet die Aussage:

$$E((b-\hat{b})^2) \geq \left(E\left(\left[\frac{\partial \ln(f(\boldsymbol{y}|b))}{\partial b} \right]^2 \right) \right)^{-1}$$
$$= -\left(E\left(\frac{\partial^2 \ln(f(\boldsymbol{y}|b))}{\partial b^2} \right) \right)^{-1}$$

- **Bemerkung:**
  - Falls ein effizienter Schätzer existiert, ist dies der ML-Schätzer.[12] [13]
  - **Übung:** Überlegen Sie sich, welche Form die Cramer-Rao-Schranke annimmt, falls die Elemente von $\boldsymbol{y}$ unabhängig und identisch verteilt sind.

[12] **Übung:** Machen Sie sich dies aufgrund des Nachweises des Theorems klar.

[13] Im Folgenden wird teilweise vorausgesetzt, dass die Entstehung des ML-Schätzers bekannt ist. „Offizielle" Definition und Herleitung erfolgt später.

Cramer-Rao-Schranke

- **Beispiel:** (Nochmal die mehrfache Messung eines Wertes)
  - Modell $\boldsymbol{y} = (1, \ldots, 1)^{\mathrm{T}} b + \boldsymbol{n}, b \in \mathbb{R}, \boldsymbol{y}, \boldsymbol{n} \in \mathbb{R}^N$
  - Ist $\boldsymbol{n}$ weiß und gaußverteilt, so folgt

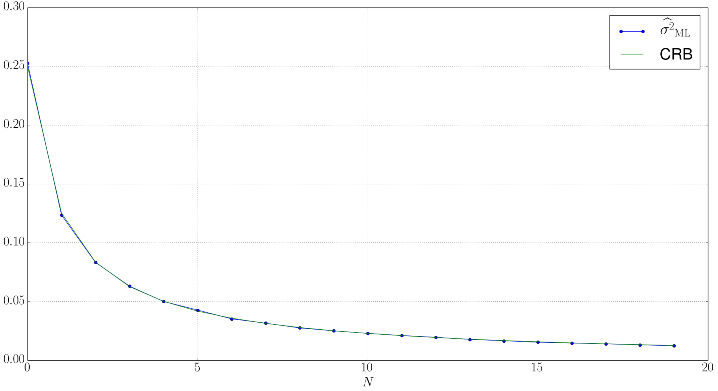
$$f(\boldsymbol{y}|b) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{\|\boldsymbol{y} - \mathbf{1}b\|^2}{2\sigma^2}\right)$$

und somit

$$J = -E\left(\frac{\partial^2 \ln(f(\boldsymbol{y}|b))}{\partial b^2}\right) = \frac{\sigma^2}{N}$$

Cramer-Rao-Schranke

- **Beispiel: (ctd.)** Cramer-Rao-Schranke und Varianz des ML-Schätzers (Mittelwertschätzer) bei 10^4 Realisierungen



Übersicht

3 Schätztheorie

- Einführung und Grundlagen
- Least-Squares-Schätzung
- Eigenschaften von Schätzern
- MMSE-Schätzer
- Verteilungsbasierte Schätzer
- Lernzielkontrolle
- Literatur

MMSE-Schätzer

- **Ziel:** Minimierung der Fehlerleistung

$$E\left(\|\boldsymbol{b}-\hat{\boldsymbol{b}}_{\mathrm{MMSE}}\|^2\right) \stackrel{!}{=} \min_{\hat{\boldsymbol{b}}} E\left(\|\boldsymbol{b}-\hat{\boldsymbol{b}}\|^2\right)$$

- Orthogonalitätsprinzip[14] ergibt:

Der Schätzfehler $\boldsymbol{b}-\boldsymbol{G}\boldsymbol{y}$ muss bei mittelwertfreiem Rauschen für den linearen MMSE-Schätzer orthogonal zu den Beobachtungen $\boldsymbol{y}$ sein. Anpassung von „orthogonal“ auf die gegebene Situation liefert:

$$E((\boldsymbol{b}-\boldsymbol{G}\boldsymbol{y})\boldsymbol{y}^{\mathrm{H}}) = \boldsymbol{0}.$$

[14] Die Tatsache folgt auch direkt durch quadratische Ergänzung und Umformung.

MMSE-Schätzer

- Nun: Lineares Signalmodell $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{n}$ mit mittelwertfreiem Rauschen, das unkorreliert von dem Parametervektor ist

MMSE-Schätzer

Theorem

Der Schätzer mit minimaler Schätzfehlerkovarianz (kurz: *Minimum-Mean-Square-Error, MMSE*), der auch als *Gauß-Markoff-Schätzer* bezeichnet wird, lautet:

$$\begin{aligned} \boldsymbol{G}_{\mathrm{MMSE}} &= \boldsymbol{R}_{\boldsymbol{bb}}\boldsymbol{X}^{\mathrm{H}}(\boldsymbol{X}\boldsymbol{R}_{\boldsymbol{bb}}\boldsymbol{X}^{\mathrm{H}} + \boldsymbol{C}_{\boldsymbol{nn}})^{-1} \\ &= (\boldsymbol{X}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{nn}}^{-1}\boldsymbol{X} + \boldsymbol{R}_{\boldsymbol{bb}}^{-1})^{-1}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{nn}}^{-1}, \end{aligned}$$

wobei[15] $\boldsymbol{R}_{\boldsymbol{bb}} = E(\boldsymbol{b}\boldsymbol{b}^{\mathrm{H}})$ und $\boldsymbol{C}_{\boldsymbol{nn}} = E(\boldsymbol{n}\boldsymbol{n}^{\mathrm{H}})$ gilt.[16].

- **Nachweis:** Tafel

[15] Es wird die Notation $\boldsymbol{R}_{\boldsymbol{bb}} = E(\boldsymbol{b}\boldsymbol{b}^{\mathrm{H}})$ verwendet, da es sich um die Korrelation und nicht um die Kovarianzmatrix handelt. Für $\boldsymbol{n}$ sind diese gleichwertig, da es mittelwertfrei ist.

[16] Für den zweiten Schritt verwende Matrix-Inversionslemma; siehe etwa [Kro91]

MMSE-Schätzer

- Bei fehlendem Vorwissen über $\boldsymbol{b}$ kann $\boldsymbol{R}_{\boldsymbol{bb}}$ nicht berechnet werden. In diesem Fall setzt man $\boldsymbol{R}_{\boldsymbol{bb}}^{-1} = \boldsymbol{0}$ und es folgt der *Minimum-Varianz-Schätzer*

$$\boldsymbol{G}_{\mathrm{MV}} = (\boldsymbol{X}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{nn}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{nn}}^{-1}$$

- **Beispiel:** ... und schon wieder die Mehrfachmessung

MMSE-Schätzer

- **Bemerkungen:**
  - Der LS-Schätzer operiert *auf einer Beobachtung* und berechnet dafür den Schätzwert mit minimalem Fehlerquadrat. Im Gegensatz dazu bestimmt der MMSE-Schätzer (GM-Schätzer) denjenigen Schätzer, der im Mittel das Fehlerquadrat minimiert.
  - Ist das Rauschen weiß und identisch verteilt (oder wird dies mangels Kenntnis angenommen), so folgt aus dem MMSE-Schätzer als Spezialfall der LS-Schätzer:

$$\boldsymbol{G}_{\mathrm{LS}} = (\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{H}}$$

    Aus [Kro91]: „[...] In diesem Fall ist die optimale Parameterschätzung identisch mit der besten Modellanpassung. [...]“. Wenn also keinerlei Wissen über Statistiken vorliegt, ist das aktuell Beste identisch mit dem im Mittel Besten.

Wiener-Filter

- **Szenario:**
  - Signal $s[k]$ wird gesendet
  - Beobachtet wird ein verrauschtes Signal: $y[k] = s[k] + n[k]$
  - Filterung der Beobachtungen mittels eines FIR-Filters mit der IA $g[k]$ liefert:

$$\hat{x}[k] = \sum_{m=0}^{N} g[m]y[k-m]$$

- **Ziel:**
  - Nicht verrauschtes Signal mit einem Filter mit der IA $w[k]$ filtern:

$$x[k] = \sum_{m=-K}^{N} w[m]s[k-m]$$

  - Gütekriterium: Quadratische Abweichung $|\hat{x}[k] - x[k]|^2$

Wiener-Filter

- **Bemerkung:** (aus [Kro91])
  - $w[n] = \delta[n]$: $x[n] = s[n]$; *Filterung* zur Beseitigung des Rauschens
  - $w[n] = \delta[n+K], K > 0$: $x[n] = s[n+K]$; *Prädiktion*
  - $w[n] = \delta[n-K], K > 0$: $x[n] = s[n-K]$; *Glättung*

  Ebenso sind auch zeitdiskrete Differenzierer, Integrierer u.v.a.m. möglich.

Wiener-Filter

Theorem (FIR-Wiener-Filter)

Sind $s[k]$ und $n[k]$ unabhängig und ist das Rauschen mittelwertfrei, so erfüllt das im Sinne der quadratischen Abweichung optimale Filter die *Wiener-Hopf-Gleichung*

$$\boldsymbol{g}^{\mathrm{T}} = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{R}_{\boldsymbol{ss}}(K)(\boldsymbol{R}_{\boldsymbol{ss}}(0) + \boldsymbol{C}_{\boldsymbol{nn}})^{-1},$$

wobei:

$$\boldsymbol{g} = (g[0], \ldots, g[N])^{\mathrm{T}} \qquad \boldsymbol{w} = (w[-K], \ldots, w[N])^{\mathrm{T}}$$

$$\boldsymbol{s}[k] = (s[k], \ldots, s[k-N])^{\mathrm{T}} \qquad \boldsymbol{n}[k] = (n[k], \ldots, n[k-N])^{\mathrm{T}}$$

$$\boldsymbol{R}_{\boldsymbol{ss}}(i) = E(\boldsymbol{s}[k+i](\boldsymbol{s}[k])^{\mathrm{T}}) \qquad \boldsymbol{C}_{\boldsymbol{nn}} = E(\boldsymbol{n}[k](\boldsymbol{n}[k])^{\mathrm{T}})$$

Wiener-Filter

- **Beispiele:** (nach [Kro91]; Entwurfsziel sei eine Filterung, d.h. $K = 0$, $w[n] = \delta[n]$)
  - Signal und Rauschen jeweils weiß und voneinander unabhängig $\Longrightarrow$ $\boldsymbol{R_{ss}}(K) = \boldsymbol{R_{ss}}(0) = \sigma_\mathrm{s}^2 \boldsymbol{I}$ und $\boldsymbol{C_{nn}} = \sigma_\mathrm{n}^2 \boldsymbol{I}$. Dann folgt:

$$g[n] = \frac{\sigma_\mathrm{s}^2}{\sigma_\mathrm{s}^2 + \sigma_\mathrm{n}^2} \delta[n] \Longrightarrow \hat{x}[n] = \frac{\sigma_\mathrm{s}^2}{\sigma_\mathrm{s}^2 + \sigma_\mathrm{n}^2} y[n].$$

  - Signal konstant, Rauschen weiß [17] $\Longrightarrow$ $\boldsymbol{R_{ss}}(K) = \boldsymbol{R_{ss}}(0) = \sigma_\mathrm{s}^2 \mathbf{1}_{N \times N}$ und $\boldsymbol{C_{nn}} = \sigma_\mathrm{n}^2 \boldsymbol{I}$. Dann folgt:

$$g[n] = \frac{1}{N + 1 + \frac{\sigma_\mathrm{n}^2}{\sigma_\mathrm{s}^2}} \mathbf{1}_{N \times 1} \Longrightarrow \hat{x}[n] = \frac{1}{N + 1 + \frac{\sigma_\mathrm{n}^2}{\sigma_\mathrm{s}^2}} \sum_{k=0}^{N} y[n-k]$$

[17] **Übung:** Zeigen Sie, dass das Ergebnis dem MMSE-Schätzer für ein konstantes Signal entspricht. Hinweis: Verwenden Sie $\boldsymbol{X} = (1, \ldots, 1)^\mathrm{T}$.

Übersicht

3 Schätztheorie

- Einführung und Grundlagen
- Least-Squares-Schätzung
- Eigenschaften von Schätzern
- MMSE-Schätzer
- Verteilungsbasierte Schätzer
- Lernzielkontrolle
- Literatur

Verteilungsbasierte Schätzer

- **Ansatz:** Es wird kein lineares Signal- oder Schätzmodell vorausgesetzt; Ausgangspunkt ist vielmehr die Verbunddichte

$$f(\boldsymbol{b}, \boldsymbol{y}).$$

- Im Weiteren werden hieraus die *a posteriori Dichte*

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

  und die *bedingte Dichte*

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

  abgeleitet und verwendet.[18]

[18] Herleitung durch die Regel von Bayes; in oben aufgeführter Formulierung wird jeweils ein Nenner ungleich Null vorausgesetzt.

MAP- und ML-Schätzer

Definition

Der *Maximum-a-posteriori (MAP)-Schätzer* ist definiert durch

$$\hat{\boldsymbol{b}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{b}} f(\boldsymbol{b}|\boldsymbol{y}).$$

Der *Maximum-Likelihood (ML)-Schätzer* erfüllt

$$\hat{\boldsymbol{b}}_{\mathrm{ML}} = \arg\max_{\boldsymbol{b}} f(\boldsymbol{y}|\boldsymbol{b}).$$

- **Folgerung:** Aufgrund der Bayes'schen Regel gilt MAP=ML, falls alle $\boldsymbol{b}$ gleichwahrscheinlich sind
- **Beispiel:**
  - Unsere gute alte Mehrfachmessung: Mittelwertschätzer ist der ML-Schätzer
  - Schätzen von Erwartungswert und Varianz einer Normalverteilung

Bayes-Schätzer

Definition

Der *Bayes-Schätzer* minimiert den Erwartungswert einer Kostenfunktion $C(\cdot)$:

$$\hat{\boldsymbol{b}}_{\text{Bayes}} = \arg\min_{\hat{\boldsymbol{b}}} E(C(\boldsymbol{b} - \hat{\boldsymbol{b}})).$$

- **Bemerkungen:**
  - Für $C(\boldsymbol{x}) = \|\boldsymbol{x}\|^2$ ergibt sich der MMSE-Schätzer.
  - Für $C(\boldsymbol{x}) = c, \boldsymbol{x} \neq 0$ entspricht dies dem MAP-Schätzer.[19]

[19] Siehe [Kro91].

MAP- und ML-Detektion

- Detektion $\rightarrow$ bekannt ist $\boldsymbol{b} \in \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_M\}$
- **Prinzip:** Aufteilen des Beobachtungsraums in *Entscheidungsregionen*:

$$\mathbb{C}^N = \bigcup_{i=1}^{M} \mathcal{R}_i, \quad \mathcal{R}_i \cap \mathcal{R}_j = \emptyset \text{ für } i \neq j$$

$$\boldsymbol{y} \in \mathcal{R}_i \Rightarrow \hat{\boldsymbol{b}} = \boldsymbol{b}_i$$

- MAP-Prinzip führt auf[20] (ML-Prinzip analog)

$$\mathcal{R}_{i,\mathrm{MAP}} = \left\{\boldsymbol{y} \in \mathbb{C}^N : f(\boldsymbol{b}_i|\boldsymbol{y}) \geq f(\boldsymbol{b}_j|\boldsymbol{y}),\ j \neq i\right\} \setminus \bigcup_{k=1}^{i-1} \mathcal{R}_{k,\mathrm{MAP}}$$

[20] **Übung:** Was bedeutet die $\setminus$-Operation, weswegen ist sie mathematisch notwendig und welche praktische Bedeutung hat sie?

MAP- und ML-Detektion

Theorem (*MAP-Theorem*)

Für die Detektionsaufgabe mit $\boldsymbol{b} \in \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_M\}$ minimiert der MAP-Schätzer die Wahrscheinlichkeit eines Detektionsfehlers ($E =$Detektionsfehler)

$$P(E) = \sum_{i=1}^{M} P(\hat{\boldsymbol{b}} \neq \boldsymbol{b}_i | \boldsymbol{b}_i) P(\boldsymbol{b}_i).$$

- **Nachweis:** Tafel

MAP- und ML-Detektion

- **Bemerkung:**
  - Sind die Symbole gleichwahrscheinlich, so minimiert der ML-Schätzer die Wahrscheinlichkeit eines Detektionsfehlers.
  - **Erinnerung/Übung:** Leiten Sie die Entscheidungsregel, die Sie aus NT I für die Demodulation kennen, mittels ML her.

Übersicht

Lernzielkontrolle

- Die folgende Aufstellung fasst die zentralen Punkte zusammen.
- Es wird aufgezeigt, welche Punkte nach Bearbeitung des Kapitels klar sein sollten.
- **Hinweise:**
  - Die Auflistung ist nicht vollständig, sondern führt die wichtigsten Aussagen auf; nicht erwähnte Inhalte sind dennoch bedeutsam.
  - Oft enthalten die Nachweise wichtige Ideen; diese also nicht vernachlässigen.
  - Stets versuchen, Gleichungen in Verbindung mit Interpretationen und Anwendungen zu sehen
  - Des weiteren sollten alle kleinen nützlichen Ergänzungen verstanden sein.
  - Es ist immer eine gute Idee, etwas Gelerntes im Rechner umzusetzen. Dies hilft beim Verständnis und schärft das Bewusstsein für mögliche Probleme.

Lernziele

Nach diesem Kapitel sollten als zentrale Punkte klar sein:

- Begriffe Schätzung und Detektion
- Lineare Schätzmodelle und lineare Schätzer, LS-Schätzer
- Bestimmung eines geeigneten Modells anhand von Beispielen
  - Schätzung eines konstanten Wertes durch Mittlung
  - Formulierung von Kanal- und Symbolschätzung bei der Nachrichtenübertragung
- Gaußapproximation und Orthogonalitätsprinzip
- Eigenschaften von Schätzern und deren Bedeutung
- Cramer-Rao-Schranke und Anwendung
- MMSE-Schätzer und Wiener-Filter; bei ersterem auch die Herleitung
- MAP- und ML-Detektion, MAP-Theorem

Übersicht

3 Schätztheorie
- Einführung und Grundlagen
- Least-Squares-Schätzung
- Eigenschaften von Schätzern
- MMSE-Schätzer
- Verteilungsbasierte Schätzer
- Lernzielkontrolle
- Literatur

Literatur I

[Kro91] H. Kronmüller, *Digitale Signalverarbeitung*, Springer, 1991