

## Schatters

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

Jansen, H. J. (1987). *Schatters: eigenschappen, constructie en overzicht*. (DCT rapporten; Vol. 1987.076). Technische Universiteit Eindhoven.

**Document status and date:**

Gepubliceerd: 01/01/1987

**Document Version:**

Uitgevers PDF, ook bekend als Version of Record

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

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

[Link to publication](#)

**General rights**

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

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

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

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

**Take down policy**

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

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

providing details and we will investigate your claim.

**Schatters: Eigenschappen,  
Constructie en Overzicht**

Verslag van H.J. Jansen

Rapport WFW87076

Afstudeerhoogleraar : Prof. Dr. Ir. J.D. Janssen  
Afstudeercoaches : Ir. W.A.M. Brekelmans  
Ir. F.J.M. Starmans

December 1987

**Vakgroep der Fundamentele Werktuigkunde  
Faculteit der Werktuigkunde  
Technische Universiteit Eindhoven**

## SYMBOLENLIJST

### Onderindex

<u>-</u>	:	kolom
o	:	waarde behorende bij vorig meetpunt
t	:	werkelijke waarde

### Bovenindex

*	:	gemeten grootheid
^	:	geschatte grootheid
T	:	getransponeerde

### Kleine letters

<u>b</u>	:	onzuiverheid	
<u>h</u>	:	relatie tussen metingen en te schatten grootheden	
k	:	stijfheid	[N/mm]
m	:	dimensie van <u>x</u>	[-]
p( <u>x</u> )	:	kansdichtheidsfunctie van <u>x</u>	
p( <u>z</u>   <u>x</u> )	:	kansdichtheidsfunctie van <u>z</u> onder de voorwaarde dat <u>x</u> is opgetreden	
r	:	verliesfunctie	
<u>y</u>	:	meetfout	
<u>x</u>	:	te schatten grootheid	
<u>z</u>	:	meting	

## Hoofdletters

A	:	weegmatrix	
E	:	verwachtingswaarde	
F	:	Fisher Informatie matrix	
H	:	<u>h</u> voor lineair model	
N of n	:	dimensie van <u>z</u>	[-]
R	:	covariantiematrix van <u>y</u>	
W	:	schatfunctie	

Als geen dimensie is aangegeven betekent dat dat de dimensie van de grootheid niet a priori bekend is.

## INLEIDING

In dit rapport wordt een overzicht gegeven van de informatie uit de literatuur over schatters.

Na een algemene beschrijving van schatproblemen wordt een aantal beoordelingscriteria van schatters besproken. Vervolgens wordt aangegeven hoe uit een bepaalde eis, bijvoorbeeld minimale variantie, een schatter voor een probleem geconstrueerd kan worden. Tenslotte wordt een overzicht gegeven van de bekendste en meest gebruikte schatters.

## 1. SCHATTEN, ALGEMEEN

### 1.1 PROBLEEMOMSCHRIJVING

Als we metingen doen, maken we altijd meetfouten. De grootte van de fout die we maken is niet a priori bekend. Daardoor is het niet zonder meer mogelijk uit metingen de werkelijke waarden van bepaalde grootheden te berekenen.

In het navolgende wordt als notatie gebruikt:

- De meetwaarden zijn geplaatst in een kolom  $\underline{z}$ . Veelal zullen we gemeten grootheden aanduiden met bovenindex \*.
- De grootheden die we willen bepalen staan in een kolom  $\underline{x}$ . De werkelijke waarden worden aangeduid met  $\underline{x}_t$ , de uiteindelijke schatting met  $\hat{\underline{x}}$ .
- De meetfouten representeren we met een kolom  $\underline{v}$ .

$$\underline{z} = \underline{h}(\underline{x}_t, \underline{v}) \quad (1.1)$$

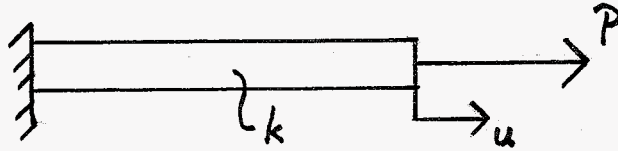
De functie  $\underline{h}$  relateert de gemeten grootheden aan de te bepalen grootheden. Over het algemeen wordt verondersteld dat de meetfouten additief gemodelleerd kunnen worden:

$$\underline{z} = \underline{h}(\underline{x}_t) + \underline{v} \quad (1.2)$$

Ofwel "de meetwaarden zijn wat we eigenlijk zouden moeten meten plus een foutje".

De kolom  $\underline{x}_t$  is onbekend (die willen we tenslotte bepalen). Ook de waarden van de grootheden in de kolom  $\underline{v}$  zijn in principe onbekend. Uiteraard is het mogelijk om bij elke waarde voor  $\underline{x}$  een kolom  $\underline{v}$  te vinden zodat exact aan (1.1) wordt voldaan.

Voorbeeld:



figuur 1.1 Trekstaaf belast door kracht P

Bepaal uitwijking u uit metingen van P en u. (bekende stijfheid k)

$$\underline{z} = [P^* \ u^*]^T \quad \underline{x} = [u] \quad \underline{h}(\underline{x}) = [ku \ u]^T$$

Stel  $k=10$ ,  $P^*=100$ ,  $u^*=10$ .

We schatten  $u=4$ , dan zou  $v_1=60$  en  $v_2=6$ .

Het zal duidelijk zijn dat hier een plausibelere schatting mogelijk is.

■

1.2 SCHATTINGSTHEORIE

Het doel van de schattingstheorie is een functie van de meetwaarden te vinden die een goede schatting levert voor de grootheden waarin we geïnteresseerd zijn.

Dus:

$$W(\underline{z}) = \hat{\underline{x}} \quad (1.3)$$

De schrijfwijze  $W(\underline{z})$  is niet helemaal correct,  $\hat{\underline{x}}$  kan ook een functie van andere grootheden, bijvoorbeeld  $\underline{y}$  zijn. De functie W wordt schatter genoemd. Het is vaak niet mogelijk een expliciete relatie te vinden als (1.3).

Uiteraard proberen we "zo goed mogelijk" te voldoen aan:

$$\hat{\underline{x}} = \underline{x}_t \quad (1.4)$$

Met andere woorden: de functie  $W(\underline{z})$  zal de invloed van de meetfouten teniet moeten doen, of in ieder geval zover mogelijk moeten reduceren.

Alvorens we de constructie van mogelijke schatters bespreken, zullen we eerst ingaan op een aantal eigenschappen, of beter: beoordelingscriteria, van schatters. Het is namelijk mogelijk om meerdere functies  $\underline{W}(\underline{z})$  te vinden, en daaruit zullen we op een of andere manier de beste moeten kiezen.

In dit rapport zal gesproken worden over de termen residu, schattingsfout en meetfout. Deze drie zijn verschillend en dienen niet met elkaar verward te worden. Daarom ter verduidelijking de volgende definities:

$$\begin{aligned} - \text{schattingsfout: } & \hat{\underline{x}} - \underline{x}_t \\ - \text{meetfout} & : \underline{z} - \underline{h}(\underline{x}_t) \\ - \text{residu} & : \underline{z} - \underline{h}(\hat{\underline{x}}) \end{aligned} \quad (1.5)$$

Hierbij wordt een perfect model  $\underline{h}(\underline{x})$  verondersteld. Dit is nooit helemaal correct.



## 2. EIGENSCHAPPEN VAN SCHATTERS

In de literatuur wordt voor de beoordeling van een schatter meestal gebruik gemaakt van de volgende vier criteria:

- 1) Zuiver (Unbiased)
- 2) Passend (Consistent)
- 3) Efficiënt (Efficient)
- 4) Voldoende (Sufficient)

Deze eigenschappen worden in dit hoofdstuk kort besproken.

### 2.1 ZUIVERHEID

De onzuiverheid van een schatter is gedefinieerd als:

$$\underline{b} = E(\hat{\underline{x}}) - \underline{x}_t \quad (2.1)$$

De verwachtingswaarde  $E$  van  $\hat{\underline{x}}$  is gedefinieerd als:

$$E(\hat{\underline{x}}) = \int \hat{\underline{x}} p(\underline{z}|\underline{x}) d\underline{z} \quad (2.2)$$

Waarbij  $p(\underline{z}|\underline{x})$  de voorwaardelijke kansdichtheidsfunctie van  $\underline{z}$  wordt genoemd. Globaal geeft deze functie de waarschijnlijkheid aan dat een bepaalde waarde  $\underline{z}$  optreedt, als bekend is dat  $\underline{x}$  is opgetreden. De verwachtingswaarde is dus een soort gewogen gemiddelde.

De verwachting is in formule (2.1) genomen over alle mogelijke realisaties van de meetwaarden.

Voorbeeld:

De N metingen  $z_1 \dots z_N$  zijn onafhankelijke samples van een Gaussische verdeling met gemiddelde  $m_t$  en variantie  $s_t^2$ .

Logische schattingen voor  $m$  en  $s^2$  zijn:

$$\hat{m} = \frac{\sum z_i}{N}$$

$$\hat{s}^2 = \frac{\sum (z_i - \hat{m})^2}{N}$$

Duidelijk is dat:  $E(\hat{m}) = m_t$ , dus de schatter voor het gemiddelde is zuiver.

Voor de verwachting van de geschatte variantie geldt:

$$E(\hat{s}^2) = \frac{1}{N} \sum \{E\{z_i - \hat{m}\}^2\} = \frac{1}{N} \sum \{E[(z_i - m) - (\hat{m} - m)]^2\} =$$
$$\frac{1}{N} \sum \{s^2 - 2s^2/N + s^2/N\} = \{(N-1)/N\} s^2$$

Hierbij is gebruik gemaakt van het feit dat  $z_1 \dots z_N$  onafhankelijke samples zijn

Dit resultaat is niet gelijk aan  $s^2$ , dus deze schatter is onzuiver. De onzuiverheid is uiteraard op te heffen door als schatter te kiezen:  $\{N/(N-1)\} \hat{s}^2$ .

Een tweede voorbeeld: Schat  $x$  uit metingen van  $\sqrt{x}$ . We veronderstellen dat de meetfout een variantie  $R$  heeft.

Dus:  $z = \sqrt{x} + v$ ,  $E\{v^2\} = R$

$$\hat{x} = z^2 \text{ (logische schatting)}$$

Nu volgt echter voor  $E\{\hat{x}\} = E\{(z+v)^2\} = x_t + R$

Ook deze schatter is niet zuiver.

■

Een zuivere schatter levert bij oneindig veel meetwaarden een schatting op die exact met de werkelijke waarde van die grootheid overeenkomt. Een asymptotisch zuivere schatter is alleen zuiver als de dimensie van de kolom met meetwaarden naar oneindig gaat.

## 2.2 PASSEND

Dit criterium geeft aan in hoeverre alle meetdata gebruikt worden. Een schatting wordt passend genoemd als:

$$\lim_{N \rightarrow \infty} \{P(|\hat{x} - x_t| < \delta)\} = 1 \quad (2.3)$$

$\delta$  = klein positief getal

$N$  = aantal meetdata

$P(y) < \delta$  : kans dat  $y < \delta$

Dus als we  $N$  maar groot genoeg maken convergeert de schatting stochastisch naar de werkelijke waarde. Dit betekent dat er best nog een realisatie van een schatter  $\hat{x}$  kan zijn, gebaseerd op een set meetwaarden  $\underline{z}$  waarvoor niet geldt dat de schatting binnen een afstand  $\delta$  van  $x_t$  ligt. Echter de kans dat dit voorkomt gaat naar nul als  $N$  naar oneindig gaat. Stochastische convergentie wil tevens zeggen dat een schatting gebaseerd op  $N+1$  meetwaarden dichter bij de werkelijke waarde van  $x$  ligt dan een schatting gebaseerd op  $N$  of minder meetwaarden. Aanvullende informatie wordt dus gebruikt.

Een passende schatting is voor een voldoende groot aantal meetwaarden zuiver. Een zuivere schatting is niet altijd passend: als we een schatting slechts baseren op de eerste twee meetdata, kan soms een zuivere schatter geformuleerd worden. Extra informatie uit eventuele verdere meetdata wordt echter niet gebruikt.

### Voorbeeld:

Gaussische verdeling als bovenomschreven.

$$\hat{m} = z_1 \quad \hat{s}^2 = \frac{1}{2}(z_1 - z_2)^2$$

zuiver, maar niet passend, daar het resultaat niet verbeterd als meer informatie beschikbaar komt. De verwachting van  $z_1$  is weliswaar  $m_t$ , maar de kans dat  $z_1$  dicht bij  $m_t$  ligt is niet afhankelijk van de dimensie van  $z$ . De kans dat de schatting dicht bij de werkelijke waarde ligt is gewoon gelijk aan de kans dat  $v_1$  heel klein is.

■

### 2.3 EFFICIENTIE

Een zuivere schatter heet efficiënter dan een andere als de variantie van de schattingsfout kleiner is.

Dus als:

$$E[\{\hat{x}_1 - x_t\}\{\hat{x}_1 - x_t\}^T] < E[\{\hat{x}_2 - x_t\}\{\hat{x}_2 - x_t\}^T] \quad (2.4)$$

Dus de efficiëntste schatter voor  $x_t$  is die zuivere schatter die de kleinste foutcovariantiematrix heeft.

In dit verband wordt met matrix A kleiner dan matrix B bedoeld dat de matrix B-A positief definitief is.

De efficiëntie van een schatter wordt vaak gemeten t.o.v. de Cramer Rao Bound. Deze bound geeft een ondergrens aan voor de haalbare foutcovariantiematrix. Hierop gaan we even nader in.

### 2.3.1 DE CRAMER RAO BOUND

De metingen met de meetfouten, alsmede de relatie tussen meet en schatgrootheden, geven een hoeveelheid informatie over de te schatten grootheden. Meer informatie zal dus niet uit de metingen te halen zijn. Met andere woorden: de nauwkeurigheid van de schatting is begrensd.

De hoeveelheid informatie kan worden weergegeven met de Fisher Informatie Matrix  $F$ . De informatie ongelijkheid stelt dat het niet mogelijk is om een schatter te construeren waarvan de covariantie

$$\text{cov}(\hat{\underline{x}}) = E\{(\hat{\underline{x}} - \underline{x}_t)(\hat{\underline{x}} - \underline{x}_t)^T\} \quad (2.5)$$

kleiner is dan de inverse Fisher Informatie Matrix.

Bovenstaande definitie van de covariantie van een schatter is enigszins afwijkend van de gebruikelijke definitie van covariantie. Daarvoor geldt namelijk:

$$\text{cov}(y) = E\{[y - E(y)][\cdot]^T\} \quad (2.6)$$

( $[\cdot]$  duidt op een herhaling van de vorige expressie tussen dezelfde haken)

Beide definities zijn alleen in overeenstemming indien een zuivere schatter gebruikt wordt. Daarvoor geldt  $E(\hat{\underline{x}}) = \underline{x}_t$ .

De matrix  $F^{-1}$  wordt de Cramer Rao Lower Bound genoemd.

Dus:

$$\text{cov}(\hat{\underline{x}}) \geq F^{-1} \quad (2.7)$$

Voor het bewijs van de informatieongelijkheid wordt verwezen naar bijlage A.

De Bound houdt in deze vorm geen rekening met eventuele schattingsfouten, er wordt uitgegaan van zuivere schatters, waarbij geen vooraf kennis aanwezig is over de te schatten grootheden. Hierop wordt onder "nadelen" teruggekomen. Voor de Fisher Informatie Matrix geldt per definitie:

$$F = E\left\{\frac{\partial}{\partial \underline{x}}[(\ln p(\underline{z}|\underline{x}))]\left[\frac{\partial}{\partial \underline{x}}(\ln p(\underline{z}|\underline{x}))\right]^T\right\} \quad (2.8)$$

Tevens geldt

$$F = -E\left\{-\frac{\partial^2}{\partial \underline{x}^2}[(\ln p(\underline{z}|\underline{x}))]\right\} \quad (2.9)$$

De mathematische achtergrond wordt hier niet behandeld.

### 2.3.2 TOEPASSING

Er zijn globaal twee toepassingen te noemen

- 1) Efficiëntie van een schatter bepalen
- 2) Optimaliseren van de metingen en/of het experiment

ad 1) De Cramer Rao Bound is onafhankelijk van de schatter. Zodra bekend is welke metingen verricht gaan worden, alsmede de relatie tussen meet- en schatgrootheden kan in principe de Bound bepaald worden. Indien de te schatten grootheid een scalar is kan de efficiëntie van de schatter die gebruikt gaat worden gedefinieerd worden als:  $F^{-1}/\text{cov}(\hat{x}(\underline{z}))$ .

Indien  $\hat{\underline{x}}$  een kolom is zal een karakteristieke matrixnorm gebruikt moeten worden. In deze vorm is de efficiëntie 1 als alle informatie uit de metingen gebruikt wordt. In elk ander geval is de efficiëntie kleiner dan 1.

ad 2) Met de Bound kan de hoogst haalbare nauwkeurigheid van het probleem berekend worden. Als een hogere nauwkeurigheid gewenst wordt zal er iets veranderd moeten worden. De invloed van nauwkeurigere metingen en/of andere metingen kan bepaald worden. In een aantal gevallen zou het overwogen kunnen worden de Bound te minimaliseren met betrekking tot een aantal nog vrij te kiezen grootheden, zoals aantal meetpunten en plaatsing van de meetpunten. Daar veranderingen vaak zowel verbeterende als verslechterende invloeden hebben op de verschillende covarianties is het aan te bevelen om een getal te gebruiken in plaats van de gehele matrix. Dat getal geeft dan globaal weer hoe nauwkeurig  $\underline{x}$  berekend zal kunnen worden.

Als voorbeelden:

$\text{tr}(AF^{-1})$ , met A een vrij te kiezen weegmatrix  
 $\text{det}(F^{-1})$ , Deze maat is objectiever aangezien de vorige maat beïnvloed wordt door een persoonlijke keuze (overigens kan dat in bepaalde gevallen juist wenselijk zijn). Aangezien een covariantiematrix een symmetrische positief-definiëte matrix is zijn alle eigenwaarden positief, en daarmee ook de determinant en het spoor.

### 2.3.3 NADELEN

Er zijn twee nadelen van deze Bound te noemen:

1) De Bound is, zoals eerder vermeld, alleen geldig voor zuivere schatters. Zeker bij niet lineaire problemen zal de schatter vaak onzuiver zijn ( $E(\hat{\underline{x}}) \neq \underline{x}_t$ ). Het is zeer wel mogelijk dat de schattingsfouten overheersen. Dat betekent dat in dat geval de Cramer Rao Bound een niet relevant minimum voor de covariantie aangeeft. In het boek van Schweppe (1973) wordt een uitbreiding gegeven van de Cramer Rao Bound die wel met schattingsfouten rekening houdt. Een nadeel is dat de Bound dan schatter afhankelijk wordt. Een ander probleem is dat zelfs als de schatter bekend is het meestal niet mogelijk is om een uitdrukking voor de schattingsfout van die schatter te geven.

2) De Bound geeft een minimum covariantie aan. Het is echter niet zeker of een schatter geconstrueerd kan worden die de minimum covariantie haalt of zelfs maar benadert. We weten dus alleen dat het geen zin heeft naar betere schatters te zoeken als de gebruikte schatter de Bound al haalt. Schweppe presenteert een verbeterde Lower Bound. In een aantal gevallen geeft die dus een hoger minimum aan dan de Cramer Rao Bound. Deze Barankin Bound is alleen toepasbaar voor zuivere schatters. Een groot nadeel voor praktisch gebruik is dat de Barankin Bound over een groot aantal vrij te kiezen grootheden gemaximaliseerd dient te worden.

Voor een aantal schatters is aangetoond dat zij onder bepaalde condities (zie onder andere de artikelen van Bar-Shalom 1972, en van Brammer 1970) de Cramer Rao Bound bereiken. Als het probleem aan die condities voldoet zal het dus niet nodig zijn naar ingenieuzere schatters te zoeken.



Voor een eerste probleemnaauwkeurighedsanalyse is een redelijk simpele formule om de covariantie van de schatter te benaderen bekend, die goed blijkt te voldoen:

Als:  $\underline{z} = \underline{h}(\underline{x}_t, \underline{v})$  en  $\hat{\underline{x}} = W(\underline{z})$  (schatter bekend !) dan geldt:

$$\text{cov}(\hat{\underline{x}}) \approx W^1 R W^{1T} \text{ met } R = E\{\underline{v}\underline{v}^T\} \quad (2.10)$$

$$\text{en } W^1 = \partial W / \partial \underline{z}$$

Dit is een gelineariseerde locale benadering waarbij geen rekening wordt gehouden met schattingsfouten. Eventueel kan de schattingsfout als volgt verdisconteerd worden :

$$\text{cov}(\hat{\underline{x}}) \approx W^1 R W^{1T} + \underline{b}\underline{b}^T \quad (2.11)$$

#### 2.4 VOLDOENDE

Een schatter heet voldoende als geen andere schatter met behulp van de meetgegevens extra informatie kan verschaffen over de te schatten grootheden.

Er zijn een aantal representaties in formulevorm, die equivalent zijn:

$$1) p(\underline{x}|\underline{z}) = p(\underline{x}|W) \quad (2.12)$$

$$2) p(\underline{x}|W, \underline{z}) = p(\underline{x}|\underline{z})$$

Deze formules zeggen dat in W dezelfde informatie zit over de kolom  $\underline{x}$  als in de kolom met meetwaarden. Dus de schatter W bevat alle informatie van de kolom  $\underline{z}$ .

Als we uit een set metingen alleen het gemiddelde nemen, is niet alle informatie gebruikt, de metingen geven ook nog een indicatie voor de nauwkeurigheid van de schatting. Het blijkt

in de praktijk vaak onmogelijk voldoende schatters te construeren.

Ideaal zou zijn een zuivere, efficiënte en voldoende schatter te hebben. Het blijkt echter vaak niet mogelijk aan alle eisen te voldoen. Veelal moet men genoegen nemen met asymptotisch zuivere en asymptotisch efficiënte schatters. Asymptotisch wil in dit verband zeggen dat er oneindig veel meetdata beschikbaar zijn (dimensie  $z$  naar oneindig). Wel dient de schatting gebaseerd te zijn op alle beschikbare data en acceptabele resultaten te geven.

De beoordeling is steeds gebaseerd op verwachtingswaarden. Voor specifieke problemen kan het soms mogelijk zijn dat bijvoorbeeld een niet efficiënte schatter beter bruikbaar is dan een efficiënte. Ook kunnen de actuele realisaties beter zijn dan de statistische verwachting. Daarom is het vaak niet mogelijk op voorhand aan te geven wat de beste schatter voor een specifiek probleem is. Wel kan worden aangegeven wat men mag verwachten van een bepaalde schatter.

Behalve op statistische eigenschappen dient een schatter tevens beoordeeld te worden op gebruiksgemak, onder andere numerieke implementeerbaarheid, rekeninspanning, benodigde geheugencapaciteit, en tuning, d.w.z. het eventueel moeten optimaliseren van bepaalde parameters in de schatfunctie.

Een zeer praktische benadering om de werking van een schatter te bekijken is de Monte Carlo Techniek:

- 1) Ga uit van een bekende  $\underline{x}_t$
- 2) Genereer (random)  $\underline{v}_j$
- 3) Bepaal  $\underline{z}_j = \underline{h}(\underline{x}_t, \underline{v}_j)$  (2.13)
- 4) Bepaal  $\hat{\underline{x}}_j = W(\underline{z}_j)$
- 5) Bepaal aan de hand van  $(\hat{\underline{x}}_j - \underline{x}_t)$   
hoe goed de schatter is.

Voordeel is dat deze methode erg simpel en algemeen toepasbaar is. Nadelen zijn dat er veel rekenwerk nodig is om een compleet beeld te krijgen van de werking van de schatter. Ook moet een goed criterium opgesteld worden om aan de hand van de  $(\hat{\underline{x}}_j - \underline{x}_t)$  de schatter te beoordelen.

### 3. ALGEMENE CONSTRUCTIE VAN SCHATTERS

In dit hoofdstuk wordt een methode besproken om schatters te construeren. Daarna zullen enkele veelgebruikte schatters besproken worden.

#### 3.1 CONSTRUCTIE VIA VERLIES FUNCTIE

Een algemene methode om een schatter, of schatfunctie te bepalen is het minimaliseren van een verlies (penalty) functie.

- 1) Kies een verliesfunctie  $L$  die aangeeft hoe zwaar een schattingsfout telt. De schattingsfout is volgens (1.5):

$$\underline{d} = \hat{\underline{x}} - \underline{x}_t. \quad (3.1)$$

Deze functie dient een grotere waarde te hebben naarmate  $\underline{d}$  groter is.

Voorbeelden:

$$(\hat{\underline{x}} - \underline{x}_t)^T (\hat{\underline{x}} - \underline{x}_t), |\hat{\underline{x}} - \underline{x}_t|, |\hat{\underline{x}} - \underline{x}_t| / \hat{\underline{x}}.$$

De eerste weegt grote fouten zwaarder (kwadratisch), terwijl de laatste een soort relatieve foutweging representeert.

- 2) Daar de waarde van  $\underline{x}_t$  niet bekend is, willen we een of andere statistische maat hebben voor het verlies als we een bepaalde schatting maken ( $\hat{\underline{x}}$  kiezen). We kiezen voor de verwachting van de verliesfunctie over alle mogelijke waarden van  $\underline{x}_t$ . We kennen de actuele waarde van  $\underline{x}_t$  niet, dus vatten we  $\underline{x}_t$  op als een variabele. We noemen deze verwachtingswaarde het gemiddelde verlies.

Uiteraard hangt de waarschijnlijkheid dat een bepaalde  $\underline{x}_t$  werkelijk is opgetreden af van de gemeten waarden  $\underline{z}$ . Het is tenslotte minder waarschijnlijk dat we een hele grote meetfout maken dan een kleine. Daarom zullen we het voorwaardelijke gemiddelde verlies bepalen. In formule:

$$r(\hat{\underline{x}}) = E\{L(\hat{\underline{x}}, \underline{x}_t) | \underline{z}\} = \int_{D\underline{x}_t} L(\hat{\underline{x}}, \underline{x}_t) p(\underline{x}_t | \underline{z}) d\underline{x}_t \quad (3.2)$$

Merk op dat  $\underline{z}$  niet als een variabele beschouwd wordt. De kolom  $\underline{z}$  heeft een actuele waarde die we kennen als we de meting gedaan hebben.

Dus als we  $\hat{\underline{x}}$  hebben genomen is het gemiddelde verlies bij de gegeven meetwaarden  $\underline{z}$  gelijk aan  $r(\hat{\underline{x}})$ . In de term  $p(\underline{x}_t | \underline{z})$  is tevens mogelijke voorkennis over de kolom  $\underline{x}_t$  verwerkt (hierop wordt nog teruggekomen onder "a posteriori schatters", in hoofdstuk 3.6).

- 3) Kies  $\hat{\underline{x}}$  zodanig dat het verlies minimaal is. Ervan uitgaande dat  $\underline{x}_t$  elke willekeurige waarde kan aannemen, kiezen we  $\hat{\underline{x}}$  zo dat ongeacht de werkelijke waarde van  $\underline{x}_t$  zijn, de kosten niet te hoog worden. Als hulp hebben we uiteraard de realisatie van de metingen:  $\underline{z}$ . We berekenen uiteindelijk:

$$\frac{dr(\hat{\underline{x}})}{d\hat{\underline{x}}} = 0 \quad (3.3)$$

Voorbeeld:

Als verliesfunctie kiezen we de lengte van de schattingsfout die we maken:  $L = (\hat{x} - x_t)^2$ . Hierbij is voor het gemak uitgegaan van een scalar  $x_t$ .

Als we deze verliesfunctie invullen in de relatie voor het voorwaardelijke gemiddelde verlies vinden we:

$$\hat{x}^2 \int_{Dx} p(x|\underline{z}) dx - 2\hat{x} \int_{Dx} xp(x|\underline{z}) dx + \int_{Dx} x^2 p(x|\underline{z}) dx$$

We willen nu  $\hat{x}$  zodanig kiezen dat dit verlies minimaal is. Daartoe differentiëren we naar  $\hat{x}$  en stellen deze afgeleide gelijk aan nul.

Eenvoudig volgt dan:

$$\hat{x} = \int_{Dx} xp(x|\underline{z}) dx$$

Per definitie geldt dat het rechterlid gelijk is aan de voorwaardelijke verwachtingswaarde van  $x$ . In het boek van Krebs wordt deze schatter verder uitgewerkt en blijkt voor dynamische systemen het Kalman Filter op te leveren.

■

Volgens de hiervoor beschreven methode is het mogelijk een aantal schatters af te leiden. In de volgende hoofdstukken presenteren we enkele veelgebruikte schatmethoden. De gekozen volgorde is van eenvoudig naar gecompliceerd. Deze presentatie komt overeen met Unbehauen (1974) en Sorenson (1980). Norton (1986) en Schweppe (1973) werken eerst het gecompliceerde, allesomvattende geval uit en laten dan steeds meer veronderstellingen vallen om de eenvoudigere methoden te presenteren. De afleiding van de methoden via de verliesfuncties blijft verder achterwege. Deze methode kan gebruikt worden als geen van de volgende schatters blijkt te voldoen.

### 3.2 KLEINSTE KWADRATEN

Indien we slechts beschikken over een model, is het toepassen van de kleinste kwadraten methode een goede keuze. Als model gebruiken we (1.2):

$$\underline{z} = \underline{h}(\underline{x}_t) + \underline{v} \quad (3.4)$$

Nu kiezen we  $\hat{\underline{x}}$  zodanig dat met alleen  $\hat{\underline{x}}$  zo goed mogelijk aan (3.4) wordt voldaan. Met andere woorden: we minimaliseren  $\underline{v}$ . Als maat gebruiken we de som van de kwadraten van de elementen van  $\underline{v}$ :  $\underline{v}^T \underline{v}$ . Daarmee:

$$L_v = (\underline{z} - \underline{h}(\underline{x}))^T (\underline{z} - \underline{h}(\underline{x})) \rightarrow \text{minimum} \quad (3.5)$$

Dit geeft de condities:

$$\frac{\partial \underline{h}(\underline{x})^T}{\partial \underline{x}} [\underline{z} - \underline{h}(\underline{x})] = 0 \quad (3.6)$$

$$\frac{\partial^2 L_v}{\partial \underline{x}^2} \text{ positief semi-definiet}$$

Dit stelsel vergelijkingen zal in het algemeen een numerieke oplossingsprocedure vergen.

Voor het geval dat (3.4) lineair is in  $\underline{x}$  kan geschreven worden:

$$\underline{z} = H \underline{x}_t + \underline{v} \quad (3.7)$$

Daarmee kan (3.6) expliciet in  $\hat{\underline{x}}$  geschreven worden. Het resultaat is:

$$\hat{\underline{x}} = (H^T H)^{-1} H^T \underline{z} \quad (3.8)$$

Zonder verdere veronderstellingen omtrent de meetfouten kunnen geen verdergaande conclusies getrokken worden.

### 3.3 GEWOGEN KLEINSTE KWADRATEN

Bij de gewogen of gegeneraliseerde kleinste kwadraten methode worden de residuen  $[z_i - h_i(\underline{x})]$  niet even zwaar gewogen. De reden hiervoor kan zijn dat men weet dat een bepaalde meting nauwkeuriger is dan andere metingen. Het is dan logisch dat aan een eventueel residu bij die meting meer gewicht wordt toegekend. In plaats van (3.5) gebruiken we nu:

$$[\underline{z} - \underline{h}(\underline{x})]^T A [\underline{z} - \underline{h}(\underline{x})] \rightarrow \text{minimum} \quad (3.9)$$

A een symmetrische, positief definitieve matrix.

Extra ten opzichte van de gewone kleinste kwadraten methode is de mogelijkheid aan bepaalde metingen extra gewicht toe te kennen, alsmede het in rekening brengen van afhankelijkheid tussen metingen.

De schatting  $\hat{\underline{x}}$  is de oplossing van

$$\frac{\partial \underline{h}(\underline{x})^T}{\partial \underline{x}} A \underline{z} = \frac{\partial \underline{h}(\underline{x})^T}{\partial \underline{x}} A \underline{h}(\underline{x}) \quad (3.10)$$

voor het niet lineaire geval. Voor het lineaire geval:

$$\hat{\underline{x}} = [H^T A H]^{-1} H^T A \underline{z} \quad (3.11)$$



### 3.4 MARKOV SCHATTING

Bij beide voorgaande methoden hebben we geen gebruik gemaakt van kennis over de meetfouten. Bij de gewogen kleinste kwadraten methode is slechts gesteld dat een meting nauwkeuriger kan zijn dan een andere, en daarom zwaarder gewogen moet worden.

Als extra informatie gaan we nu statistische eigenschappen van de kolom  $\underline{y}$  introduceren. Veronderstel dat  $E\{\underline{y}\} = 0$  en  $E\{\underline{y} \underline{y}^T\} = R$ .

Met deze extra kennis kunnen we iets zeggen over de kwaliteit van de schattingen. We beperken ons tot het lineaire geval met additieve meetfouten volgens (1.2):  $\underline{z} = H\underline{x} + \underline{y}$

De schatter wordt gezocht in de klasse van lineaire schatfuncties volgens:

$$\hat{\underline{x}} = A\underline{z} + A_0 \quad (3.12)$$

En dus:

$$E(\hat{\underline{x}}) = E\{A(\underline{z}) + A_0\} = E\{AH\underline{x}_t + A\underline{y} + A_0\} \quad (3.13)$$

Voor een zuivere schatter moet gelden:

$$AH = I \text{ en } A_0 = 0 \quad (3.14)$$

omdat  $A$  en  $A_0$  niet van  $\underline{x}$  mogen afhangen.

We bepalen nu de covariantiematrix van de schatting:

$$E\{(\hat{\underline{x}} - \underline{x}_t)(\hat{\underline{x}} - \underline{x}_t)^T\} = E\{(A\underline{z} - \underline{x}_t)(A\underline{z} - \underline{x}_t)^T\} = E\{(AH\underline{x}_t + A\underline{y} - \underline{x}_t)(AH\underline{x}_t + A\underline{y} - \underline{x}_t)^T\} = ARA^T \quad (3.15)$$

Voor minimale variantie bepalen we het minimum van (3.15) onder de nevenvoorwaarde (3.14):

$$ARA^T + (AH - I)C \rightarrow \text{minimum} \quad (3.16)$$

C is een matrix met Lagrange Multiplicatoren.  
Oplossen van 3.16 levert:

$$A = [H^T R^{-1} H] H^T R^{-1} \quad (3.17)$$

Zodat volgt:

$$\hat{\underline{x}} = [H^T R^{-1} H]^{-1} H^T R^{-1} \underline{z} \quad (3.18)$$

Deze schatter is zuiver en heeft van alle lineaire zuivere schatters de laagste covariantie. We zien dat deze zogeheten Markov-schatter overeenkomt met de gewogen kleinste kwadraten schatter, indien voor de weegmatrix de inverse covariantiematrix van de meetfouten gekozen wordt.

Dit betekent:

- 1) Gewone kleinste kwadraten (hoofdstuk 3.2) heeft een minimale covariantie indien  $R = s^2 I$ , met  $s^2$  de variantie van de meetfouten.
- 2) Gewogene kleinste kwadraten (hoofdstuk 3.3) heeft een minimale covariantie indien weegmatrix  $A = R^{-1}$

Opgemerkt wordt dat het soms mogelijk is een onzuivere schatter te vinden met een lagere covariantie. Deze schatters worden behandeld in het boek van Norton onder het hoofdstuk Ridge Regression.

### 3.5 MAXIMUM LIKELIHOOD

Indien we over nog meer kennis van de meetruis beschikken kunnen we een in het algemeen betere schatter afleiden: we kiezen de schatting  $\hat{\underline{x}}$  zodanig dat  $p(\underline{z}|\underline{x})$  gemaximaliseerd wordt. De voorwaardelijke kansdichtheidsfunctie  $p(\underline{z}|\underline{x})$  kan opgesteld worden zodra de modelbeschrijving en de kansdichtheidsfunctie van de meetruis beschikbaar zijn. De voorwaardelijke kansdichtheidsfunctie  $p(\underline{z}|\underline{x})$  is, zodra we voor  $\underline{z}$  actuele waarden kunnen invullen nog slechts een functie van  $\underline{x}$ . "We kiezen  $\hat{\underline{x}}$  dan zodat het heel waarschijnlijk is dat de gemaakte meting bij die  $\underline{x}$  hoort".

$$\hat{\underline{x}} \text{ is oplossing van } \frac{\partial p(\underline{z}|\underline{x})}{\partial \underline{x}} = 0 \quad (3.19)$$

Voor niet lineaire schatproblemen is deze schatter asymptotisch zuiver, passend en asymptotisch efficiënt. Deze eigenschappen gelden in ieder geval als de metingen onafhankelijk zijn. Ingeval van afhankelijke metingen gelden de eigenschappen onder bepaalde condities (Brammer 1970). Tevens is aan te tonen dat indien de Cramer Rao Bound bereikt kan worden, dit mogelijk is met de Maximum Likelihood Schatter (bijlage B, Sorensson 1980).

### 3.6 MAXIMUM A POSTERIORI ESTIMATORS

In een aantal gevallen hebben we bij het schatten vooraf al kennis over de te schatten grootheden. Als we bijvoorbeeld de stijfheidsmodulus van een stalen staaf uit metingen willen bepalen, kunnen we van te voren al vertellen dat deze ongeveer  $2.1 \cdot 10^5 \text{ N/mm}^2$  is.

Met behulp van deze kennis is het mogelijk de schatting ten opzichte van de maximum likelihood methode te verbeteren. We gaan ervan uit dat we van de kolom met te schatten grootheden kunnen zeggen:

$$E\{\underline{x}\} = \underline{m} \quad (3.20)$$

$$E\{(\underline{x} - \underline{m})(\underline{x} - \underline{m})^T\} = Y \quad (3.21)$$

Verder nemen we als model (1.2), met extra veronderstellingen over de meetfouten

$$\begin{aligned} \underline{z} &= H\underline{x} + \underline{v} & E\{\underline{v}\} &= 0 \\ & & E\{\underline{v}\underline{v}^T\} &= R \end{aligned} \quad (3.22)$$

Dan levert eenzelfde afleiding als gegeven onder de Markov schatters (hoofdstuk 3.4):

$$\hat{\underline{x}} = \underline{m} + YH^T[HYH^T+R]^{-1}[\underline{z}-H\underline{m}] \quad (3.23)$$

met covariantie:

$$\text{cov}(\hat{\underline{x}}) = Y - YH^T[HYH^T+R]^{-1}HY \quad (3.24)$$

Deze schatter is zuiver en heeft voor dit geval de laagste covariantie van alle zuivere lineaire schatters.

In het algemene, niet noodzakelijkerwijs lineaire, geval waarbij we beschikken over kansverdelingsfuncties voor  $\underline{v}$  en  $\underline{x}$  in plaats van slechts hun gemiddelde en covariantiematrix, wordt de MAP (maximum a Posteriori) schatting gebruikt:

$$\text{Maximaliseer } p(\underline{x}|\underline{z}) \quad (3.25)$$

Volgens Bayes Rule geldt:

$$p(\underline{x}|\underline{z}) = p(\underline{z}|\underline{x}) \frac{p(\underline{x})}{p(\underline{z})} \quad (3.26)$$

Ten opzichte van de Maximum Likelihood schatter zien we de voorkennis over de parameters in de term  $p(\underline{x})$ . Vaak is het erg moeilijk en bewerkelijk de hele MAP schatting uit te werken. Opgemerkt wordt nog dat als alle realisaties van  $\underline{x}$  even waarschijnlijk zijn, m.a.w.  $p(\underline{x}) = c$ , dan levert maximalisatie van  $p(\underline{z}|\underline{x})$  hetzelfde op als maximalisatie van  $p(\underline{x}|\underline{z})$ . Schweppe (1973) en Sorenson (1980) merken verder op dat de voorkennis uit de MAP schatting ook in de Maximum Likelihood Formulering kan worden ingebracht via een zogenaamde extra meting:

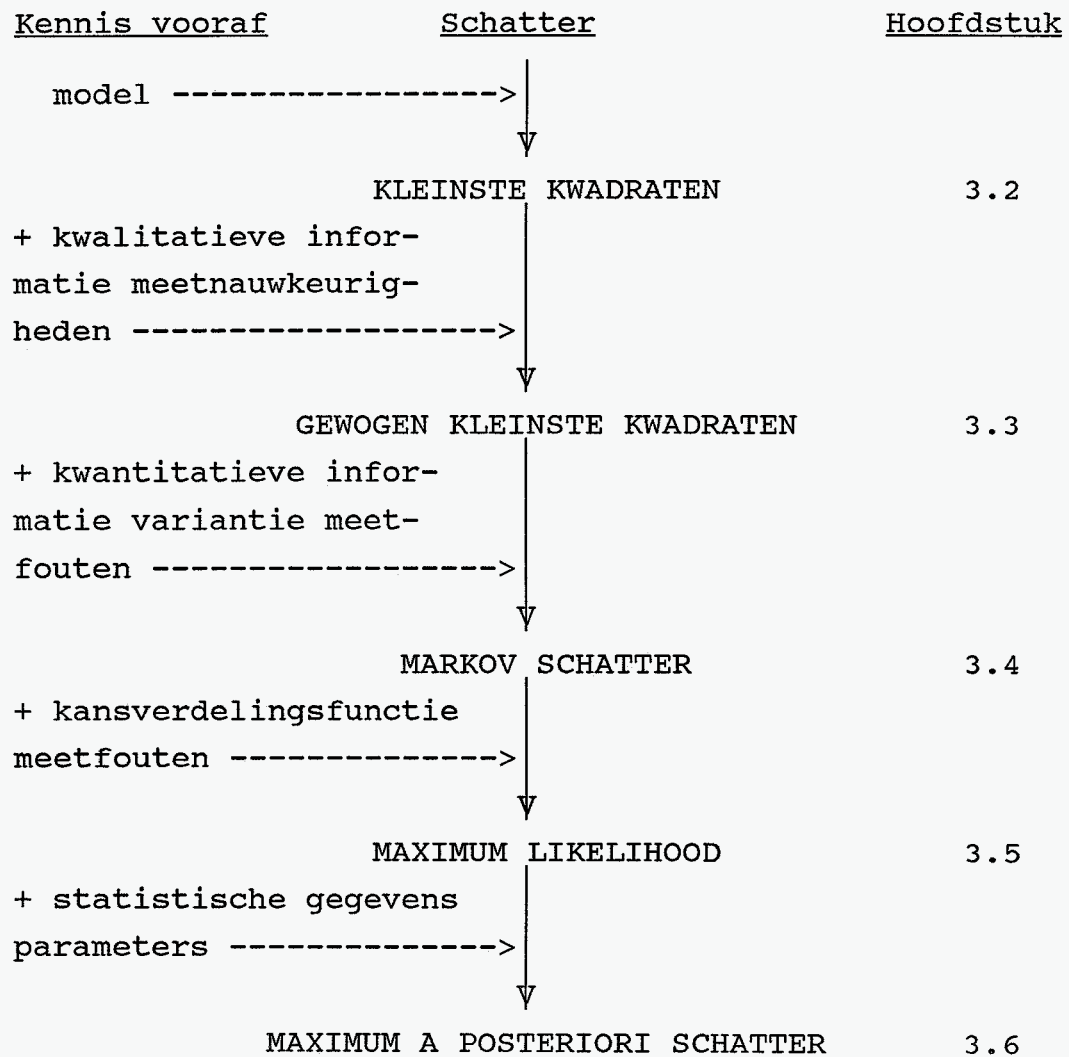
$\underline{z}_{extra} = \underline{x} + \underline{v}_{extra}$ , waarbij voor  $\underline{z}$  de waarde  $\underline{m}$  wordt ingevuld, en voor de covariantie van de betreffende meetfout  $\underline{v}$  de matrix  $Y$ . De covariantie van deze meetfout met de overige is gelijk nul.

(3.27)

Overigens is het ook voor het niet-lineaire geval mogelijk een minimum variantie schatting uit te werken (Krebs (1980)). Dit is echter zeer bewerkelijk.

### 3.7 BESLUIT

In dit hoofdstuk zijn de bekendste schatmethoden kort besproken. Vanuit weinig kennis over de verschillende grootheden, hebben we steeds meer toegevoegd. Onderstaand schema vat een en ander nog eens samen.



Dit schema is zeker niet compleet. Er zijn vanuit de gepresenteerde verliesfunctiemethode nog vele andere schatters af te leiden. Hiervoor wordt verder verwezen naar Norton (1986), Unbehauen (1974), en Sorenson (1980).

De formulering van het schatprobleem met voorkennis over de parameters staat bekend als Bayes Formulering, de formulering zonder voorkennis over de parameters als Fisher Formulering. Over het algemeen levert de Bayes schatting betere resultaten, mits de voorkennis ( $p(\underline{x})$ ) goed gemodelleerd is. Het is echter gevaarlijk om voorkennis die niet goed is, te modelleren. Dit werkt namelijk vaak averechts en men kan dan beter een schatter gebruiken die minder voorkennis vereist. Tevens wordt er door Sorenson (1980) op gewezen dat de verbetering die men bereikt met de Bayes ten opzichte van de Fisher Formulering voornamelijk tot uiting komt bij kleine sets meetwaarden (dimensie  $z$  laag). Bij grote sets meetwaarden is de verbetering zeer gering terwijl vaak een veel gecompliceerder schatalgorithme gebruikt moet worden. Unbehauen (1974) komt tot de conclusie dat Maximum Likelihood over het algemeen goed voldoet (niet lineair probleem), terwijl de rekeninspanning laag blijft. Als grootste probleem noemt hij de grote geheugencapaciteit die nodig is.

In hoofdstuk 4 besteden we aandacht aan het recursief verwerken van meetdata. Dit kan nodig zijn om het geheugengebruik te beperken bij toepassing van bijvoorbeeld de Maximum Likelihood Methode.

#### 4. HET RECURSIEF PARAMETER SCHATTEN

Als we een aantal opeenvolgende metingen doen aan een parameter of een toestand kan het wenselijk zijn deze on-line (achtereenvolgens) te verwerken. Dit kan vanwege de wens onmiddellijk informatie te verkrijgen uit een meting dan wel vanwege de beschikbare computercapaciteit die verwerking van meerdere metingen tegelijk niet toelaat.

Indien gebruik gemaakt wordt van de recursieve methode kan een parallel getrokken worden met het Kalman Filter.

Er zijn twee mogelijke wegen om tot een dergelijke formulering te komen.

##### 4.1 RECURSIEVE FORMULERING

Na verwerking van N metingen hebben we een schatting voor de parameters  $\hat{\underline{x}}$  die we aanduiden met  $\hat{\underline{x}}_N$ . Om de volgende meting te verwerken nemen we aan dat alle informatie uit de voorgaande metingen verwerkt is in  $\hat{\underline{x}}_N$  en de bij  $\hat{\underline{x}}_N$  behorende covariantiematrix  $P_N$ . We bepalen dan de schatting gebaseerd op N+1 metingen uit de schatting  $\hat{\underline{x}}_N$  en meting  $\underline{z}_{N+1}$ . We zullen dus een goede weging moeten vinden om de informatie uit de meting te combineren met de informatie die we al hebben op grond van de vorige metingen. Deze vraagstelling komt overeen met de vraagstelling uit hoofdstuk 3.6, de Bayes Formulering. We gaan uit van een lineair schattingsprobleem:

$$\underline{z}_N = H_N \underline{x}_t + \underline{v} \quad (4.1)$$

$$E\{\underline{v}\} = 0$$

$$E\{\underline{v}\underline{v}^T\} = R$$



Verder nemen we aan dat we een zuivere schatter voor de kolom  $\underline{x}$  hebben, dus:

$$E\{\hat{\underline{x}}\} = \underline{x}_t \quad (4.2)$$

We zullen nu een schatter afleiden gebaseerd op N+1 metingen en deze aanduiden met  $\hat{\underline{x}}_{N+1}$ . Deze wordt gekozen als zijnde lineair afhankelijk van  $\hat{\underline{x}}_N$  en van  $\underline{z}_{N+1}$  (deze keuze beperkt uiteraard de algemeenheid).

In formule:

$$\hat{\underline{x}}_{N+1} = J_{N+1}\hat{\underline{x}}_N + K_{N+1}\underline{z}_{N+1} \quad (4.3)$$

Deze formule is te interpreteren als een weging van vorige metingen tegenover de huidige. Om aan de eis van een zuivere schatter te voldoen moet gelden:

$$J_{N+1} + K_{N+1}H_{N+1} = I \quad (4.4)$$

Met deze vergelijking elimineren we  $J_{N+1}$ , waarna volgt:

$$\hat{\underline{x}}_{N+1} = \hat{\underline{x}}_N + K_{N+1}\{\underline{z}_{N+1} - H_{N+1}\hat{\underline{x}}_N\} \quad (4.5)$$

In (4.5) is  $K_{N+1}$  te bepalen. Daartoe eisen we een minimum covariantie voor de schatting  $\hat{\underline{x}}_{N+1}$ . De vergelijking voor de covariantiematrix  $P_{N+1}$  luidt:

$$P_{N+1} = \{I - K_{N+1}H_{N+1}\}P_N\{I - K_{N+1}H_{N+1}\}^T + K_{N+1}R K_{N+1}^T \quad (4.6)$$

(Formule (4.6) is rechtstreeks af te leiden uit  $\text{cov}(\hat{\underline{x}}) = E\{(\hat{\underline{x}}_{N+1} - \underline{x}_t)(\hat{\underline{x}}_{N+1} - \underline{x}_t)^T$ .)

We bepalen nu hoe de covariantiematrix verandert als gevolg van kleine veranderingen in de matrix  $K_{N+1}$ . Variatie van  $K_{N+1}$  levert variatie van  $P_{N+1}$ . Ter bepaling van de minimum covariantie eisen we dat de variatie van de covariantiematrix nul is voor willekeurige variatie van  $K_{N+1}$ . Voor  $K_{N+1}$  volgt dan de volgende vergelijking:

$$-(I - K_{N+1}H_{N+1})P_N H_{N+1}^T + K_{N+1}R = 0 \quad (4.7)$$

En daarmee expliciet:

$$K_{N+1} = P_N H_{N+1}^T (H_{N+1} P_N H_{N+1}^T + R)^{-1} \quad (4.8)$$

$$P_{N+1} = \{I - K_{N+1}H_{N+1}\}P_N$$

Bij de hierboven beschreven strategie wordt na verloop van tijd aan nieuwe meetwaarden geen enkel gewicht meer toegekend omdat de schatting van  $\underline{x}_t$  nauwkeuriger aangenomen wordt dan de meting. Om toch steeds de informatie uit een meting te kunnen verwerken zouden we de covariantiematrix  $P$  af en toe kunnen instellen, hetgeen het toekennen van een fictieve nauwkeurigheid aan de schatting inhoudt. De meting direct na de instelling wordt dan echter onevenredig zwaar meegenomen zonder dat daar een goede reden voor is. We zullen een andere methode presenteren die een soort systeemruis introduceert. Daarmee stellen we eigenlijk dat we de parameters nooit precies kennen. Feitelijk staan we zelfs toe dat ze heel langzaam variëren in de tijd. Een goede fysische achtergrond is hier niet voor te geven. Er zou gedacht kunnen worden aan het verwerken van modelonnauwkeurigheid in de systeemruis. De systeemruis geeft dan aan in hoeverre de werkelijkheid afwijkt van het gebruikte model.

## 4.2 RECURSIEVE FORMULERING MET SYSTEEMRUIS

Om te voorkomen dat de versterkingsmatrix  $K_N$  na verloop van tijd te klein wordt, introduceren we een soort systeemruis. Voor het schatten van parameters doen we dit als volgt:

$$\underline{x}_{N+1} = \underline{x}_N + \underline{w}_N \text{ (systeemvergelijking)} \quad (4.9)$$

$$E\{\underline{w}_N\} = 0$$

$$\text{COV}\{\underline{w}_N\} = Q$$

Definiëer nu:

$$\begin{aligned} \hat{\underline{x}}_{N+1|N} &= \hat{\underline{x}}_N \\ P_{N+1|N} &= P_N + Q \end{aligned} \quad (4.10)$$

Met andere woorden, de vooraf kennis is de schatting gebaseerd op de N reeds verwerkte metingen. Gebruik dit als beschreven in hoofdstuk 4.1. Deze methode levert exact de Kalman Filter vergelijkingen.

## 4.3 CONCLUSIES

Het Kalman Filter is gebaseerd op een Bayes Schattings Formulering waarbij als schatter de minimum covariantie schatting gebruikt wordt. Een uitgebreide afleiding, ook voor het niet lineaire geval, is te vinden in het boek van Volker Krebs. Bij het parameterschatten kunnen we een dergelijke formulering verkrijgen door de metingen achtereenvolgens te verwerken. Voordelen van deze werkwijze zijn dat minder computercapaciteit nodig is daar het op te lossen stelsel kleiner is en dat voor het

verwerken van de Kalman Filter vergelijkingen vele algoritmen beschikbaar zijn. Nadeel is dat voor elke meting (eventueel set metingen) het stelsel vergelijkingen opnieuw moet worden opgelost. Bij niet lineaire problemen houdt dat in dat een minimalisatie procedure vele malen doorlopen moet worden. Als alle metingen in een keer verwerkt worden hoeft dit slechts een maal te geschieden. Echter met een groter stelsel vergelijkingen. Hier zal proberen uitsluitel moeten geven over welke werkwijze het effectiefst is.

## 5. CONCLUSIES

Voor het schatten van toestanden en/of parameters (het verschil is niet geheel duidelijk), zijn een aantal methoden bekend in de literatuur. Afhankelijk van de aard van het model (lineair of niet lineair) en de beschikbare a priori kennis (meetfoutvarianties en dergelijke) kan een van de methoden gekozen worden. De keuze wordt tevens bepaald door geheugen- en rekencapaciteit van de beschikbare computerfaciliteiten.

Het overzicht in hoofdstuk 3.7 beoogt een leidraad te zijn om tot een keuze te komen. De gegeven behandeling is echter verre van compleet en dient niet als bindend beschouwd te worden.

## LITERATUUR

### Boeken

Krebs, V.

Nichtlineare Filterung

Oldenburg Verlag 1980

Norton, J.P.

An Introduction to Identification

Academic Press 1986

Sage and Melsa

System Identification

Academic Press 1971

Schweppe, F.C.

Uncertain Dynamic Systems

Prentice Hall 1973

Sorenson, H.W.

Parameter Estimation

Control and Systems Theory vol. 9 1980

Unbehauen, H. Goehring, B. Bauer, B.

Parameterschätzverfahren zur Systemidentifikation

Oldenbourg Verlag 1974

## Artikelen

Bar-Shalom, Y.

Parameter Identification in Linear Discrete-Time Systems  
IEEE Transactions on Automatic Control, vol. AC-17 1972

Brammer, K.

Schätzung von Parametern und Zustandsvariablen linearer  
Regelstrecken durch nichtlineare Filterung  
Regelungstechnik und Prozeßdaten Verarbeitung Heft 6 1970

## TUE-Literatuur

Bosch, A.J. Kamps, H.J.L.

Statistisch Compendium

Dictaat Kansrekening en Statistiek

Kok, J.J.

Dictaat Werktuigkundig Regelen II

Hendriks, M.

Bepaling van de Materiaaleigenschappen van Biologische  
Materialen met behulp van Systeemidentificatietechnieken  
Afstudeerverslag 1986

Linssen, F.M.J.

Bepaling van de Materiaaleigenschappen met behulp van  
Filtertechnieken

Afstudeerverslag 1987

Starmans, F.J.M.

Toepassing van het Statisch Lineaire Schattingsmodel volgens  
Fisher op het Lineair Elasticiteitsprobleem met een  
Gevarieerde Set van Randvoorwaarden

WFW-Rapport 1986



## INHOUDSOPGAVE

### SYMBOLENLIJST

INLEIDING	1
1. SCHATTEN, ALGEMEEN	2
1.1 PROBLEEMOMSCHRIJVING	2
1.2 SCHATTINGSTHEORIE	3
2. EIGENSCHAPPEN VAN SCHATTERS	5
2.1 ZUIVERHEID	5
2.2 PASSEND	7
2.3 EFFICIENTIE	8
2.3.1 DE CRAMER RAO BOUND	9
2.3.2 TOEPASSING	10
2.3.3 NADELEN	12
2.4 VOLDOENDE	13
3. ALGEMENE CONSTRUCTIE VAN SCHATTERS	16
3.1 CONSTRUCTIE VIA VERLIES FUNCTIE	16
3.2 KLEINSTE KWADRATEN	19
3.3 GEWOGEN KLEINSTE KWADRATEN	20
3.4 MARKOV SCHATTING	21
3.5 MAXIMUM LIKELIHOOD	23
3.6 MAXIMUM A POSTERIORI ESTIMATORS	23
3.7 BESLUIT	26
4. HET RECURSIEF PARAMETER SCHATTEN	28
4.1 RECURSIEVE FORMULERING	28
4.2 RECURSIEVE FORMULERING MET SYSTEEMRUIS	31
4.3 CONCLUSIES	31
5. CONCLUSIES	33
LITERATUUR	34

## APPENDIX A: DE CRAMER RAO BOUND

Als we een hoeveelheid metingen doen, hebben we een hoeveelheid informatie over de grootheden die we willen bepalen. De nauwkeurigheid waarmee we grootheden kunnen schatten is onder andere door die hoeveelheid informatie begrensd. Hoe meer informatie, hoe nauwkeuriger de schatting. De hoeveelheid informatie die de metingen bevatten kunnen we aangeven met de Fisher Informatie Matrix  $F$ . De matrix  $F$  is per definitie gelijk aan:

$$F = E \left[ \left[ -\frac{\partial}{\partial \underline{x}} \ln\{p(\underline{z}|\underline{x})\} \right] \left[ -\frac{\partial}{\partial \underline{x}} \ln\{p(\underline{z}|\underline{x})\} \right]^T \right] \quad (A1)$$

Tevens is aan te tonen dat:

$$F = -E \left[ \frac{\partial^2}{\partial^2 \underline{x}} \ln\{p(\underline{z}|\underline{x})\} \right] \quad (A2)$$

In het bovenstaande is ervan uitgegaan dat de kolom  $\underline{x}$  volledig onbekend is.

### Bewijs Informatieongelijkheid

Beschouw de kolom

$$\underline{\phi} = \{ \hat{\underline{x}}(\underline{z}) \quad -\frac{\partial}{\partial \underline{x}} (\ln p(\underline{z}|\underline{x})) \}^T \quad (A3)$$

De kolom  $\underline{x}$  bevat de te schatten grootheden.

We gaan nu de covariantie van  $\underline{\phi}$  bepalen,

$$\text{cov}(\underline{\phi}) = E\{ \underline{\phi} - E(\underline{\phi}) \} \{ \cdot \}^T \quad (A4)$$

De punt tussen de haakjes duidt op een herhaling van de voorgaande expressie tussen haakjes.

$$1a) \quad E(\hat{\underline{x}}) = \underline{x} \quad (A5)$$

omdat we uitgaan van een zuivere schatter

$$1b) \quad E\left\{-\frac{\partial}{\partial \underline{x}}(\ln p(\underline{z}|\underline{x}))\right\} = \int_{D_z} \left(-\frac{\partial}{\partial \underline{x}}(\ln p(\underline{z}|\underline{x}))\right) p(\underline{z}|\underline{x}) d\underline{z} = \quad (A6)$$

$$= -\frac{\partial}{\partial \underline{x}} \left( \int_{D_z} p(\underline{z}|\underline{x}) d\underline{z} \right) = 0$$

Deze afleiding stelt netheidseisen aan de voorwaardelijke kansdichtheidsfunctie zodanig dat de integratie en differentiatie bewerkingen in volgorde mogen worden omgewisseld.

We gaan nu (A4) bepalen:

$$2a) \quad E\{[\hat{\underline{x}} - E(\hat{\underline{x}})][\cdot]^T\} = \text{cov}(\hat{\underline{x}}) \quad (A7)$$

$$2b) \quad E\left\{\hat{\underline{x}}(\underline{z}) * \frac{\partial}{\partial \underline{x}}(\ln p(\underline{z}|\underline{x})) - E(\hat{\underline{x}})E\left[\frac{\partial}{\partial \underline{x}}(\ln p(\underline{z}|\underline{x}))\right]\right\} =$$

$$= \int_{D_z} \left\{\hat{\underline{x}}(\underline{z}) * \frac{\partial}{\partial \underline{x}}(\ln p(\underline{z}|\underline{x})) p(\underline{z}|\underline{x})\right\} d\underline{z} = -\frac{\partial}{\partial \underline{x}} \left\{ \int_{D_z} [\hat{\underline{x}}(\underline{z}) p(\underline{z}|\underline{x})] d\underline{z} \right\} =$$

$$= \frac{\partial}{\partial \underline{x}}(\underline{x}) = I \quad (A8)$$

$$2c) \quad E\left\{\left[-\frac{\partial}{\partial \underline{x}}(\ln p(\underline{z}|\underline{x})) - E\left\{-\frac{\partial}{\partial \underline{x}}(\ln p(\underline{z}|\underline{x}))\right\}\right][\cdot]^T\right\} = F \quad (A9)$$

Vergelijkingen A5 tot en met A9, gesubstitueerd in vergelijking A4 levert voor de covariantiematrix van  $\varnothing$ :

$$\text{cov}(\varnothing) = \begin{bmatrix} \text{cov}\{\hat{\underline{x}}(\underline{z})\} & \text{I} \\ \text{I} & \text{F} \end{bmatrix} \quad (\text{A10})$$

De covariantiematrix is per definitie positief semi-definiet zodat:

$$\underline{a}^T \begin{bmatrix} \text{I} & \text{F}^{-1} \\ \text{I} & \text{F} \end{bmatrix} \begin{bmatrix} \text{cov}\{\hat{\underline{x}}(\underline{z})\} & \text{I} \\ \text{I} & \text{F} \end{bmatrix} \begin{bmatrix} \text{I} \\ \text{F}^{-1} \end{bmatrix} \underline{a} > 0 \quad (\text{A11})$$

Formule (A9) geldt voor willekeurige kolom  $\underline{a}$ . Uitwerking levert:

$$\begin{bmatrix} \text{cov}\{\hat{\underline{x}}(\underline{z})\} - \text{F}^{-1} \end{bmatrix} \geq 0 \quad (\text{A12})$$

Hiermee is de informatieongelijkheid voor zuivere schatters bewezen.

Er is een aantal extra eisen aan de kansdichtheidsfunctie  $p(\underline{z}|\underline{x})$  nodig. Uitwerking hiervan is te vinden in het artikel van Brammer (1972).

De informatieongelijkheid is in appendix A afgeleid met de veronderstelling dat  $\underline{x}$  een kolom met volledig onbekende grootheden is. Indien  $\underline{x}$  een stochastische grootheid is met bekende kansdichtheidsfunctie  $p(\underline{x})$ , dan dient in de informatieongelijkheid uit appendix A de Fisher Informatiematrix  $F$  vervangen te worden door  $L$ . Voor de matrix  $L$  geldt (Sorenson 1980):

$$L = E\left\{\left[-\frac{\partial}{\partial \underline{x}}(\ln p(\underline{x}, \underline{z}))\right][\cdot]^T\right\} \quad (B4)$$

In dit geval kan gelijkheid alleen bereikt worden indien:

$$\left[-\frac{\partial}{\partial \underline{x}}(\ln p(\underline{x}, \underline{z}))\right]^T = C[\underline{x} - \hat{\underline{x}}] \quad (B5)$$

In (B5) is  $C$  onafhankelijk van  $\underline{x}$  en  $\underline{z}$ . Hieraan kan alleen voldaan worden als  $p(\underline{x}, \underline{z})$  Gaussisch is (Sorenson 1980).

Uit bovenstaande verhandeling blijkt dat het zeker niet algemeen mogelijk is de Cramer Rao Bound te bereiken. Voor niet lineaire schatproblemen is meestal geen analytische uitdrukking te geven voor de covariantiematrix van de schattingsfout of voor de schattingsfout zelf.