Toegepaste Wiskunde 2: Het Kalman-filter

Hans Maassen

25 februari, 2008

1. Inleiding

Het Kalman filter schat de toestand van een systeem op basis van een reeks, door ruis verstoorde waarnemingen. Een meer correcte benaming zou zijn: het lineaire Stratonovich-Kalman-Bucy filter.
Het is zeer geschikt om electronica te programmeren die vliegende objecten in *real time* localiseert. Tegenwoordig wordt het filter ook ingezet om beurskoeren te voorspellen, en diverse andere, in de tijd veranderende grootheden.
Het filter wordt gevoed met stochastische *a priori* aannamen over bijvoorbeeld plaats en snelheid van het te volgen object, en vertaalt deze na elke nieuwe waarneming in bijgestelde, *a posteriori* schattingen. De schatting op zeker tijdstip is de *voorwaardelijke verwachting* van de betreffende grootheid, gegeven alle observaties die tot dan toe zijn gedaan.
De geschiedenis van het Kalmanfilter is een van de succesverhalen van de Toegepaste Wiskunde. Meteen na de ontwikkeling ervan door Kalman, op basis van ideeën van Stratonovich, is het door de NASA ingezet in het Apolloproject, dat ten doel had een mens op de maan te brengen. “Without me no moon”, placht Kalman te zeggen.

2. Voorwaardelijke Verwachting

De *voorwaardelijke kans* op een gebeurtenis A, gegeven een gebeurtenis B is gedefinieerd als

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} ,$$

mits $\mathbb{P}(B) \neq 0$. Dit leidt onmiddellijk tot de definitie van een *voorwaardelijke verwachting* van een discrete stochast X, gegeven een gebeurtenis B: zij $\mathcal{W}$ de waardenverzameling van X, dan is

$$\mathbb{E}(X|B) := \sum_{x \in \mathcal{W}} x \, \mathbb{P}([X = x]|B) .$$

En deze leidt weer tot de definitie van een *voorwaardelijke verwachting* van een discrete stochast X, gegeven een tweede discrete stochast Z:

$$\mathbb{E}(X|Z) : \Omega \to \mathbb{R} : \omega \mapsto f(Z(\omega)) \ , \ \text{waarbij } f(z) := \mathbb{E}(X|Z = z) \ . \tag{1}$$

Merk op dat deze voorwaardelijke verwachting zelf een stochast is, helemaal vastgelegd door Z. We beweren nu dat dit de beste schatting is van X op basis van kennis over Z.

Lemma 1. *Laten X en Z discrete stochasten zijn, zeg met waardeverzamelingen $\mathcal{W}_X$ en $\mathcal{W}_Z$. Dan is $\mathbb{E}(X|Z) = f(Z)$, waarbij f díe functie is van $\mathcal{W}_Z$ naar $\mathbb{R}$, waarvoor*

$$\mathbb{E}\big((f(Z) - X)^2\big)$$

minimaal is.

Dit is een uitbreiding van het eenvoudige stellinkje dat zegt dat voor een willekeurige stochast X de waarde van $\lambda \in \mathbb{R}$ waarvoor $\mathbb{E}\big((\lambda - X)^2\big)$ minimaal is gegeven wordt door de verwachting $\mathbb{E}(X)$. Immers: voor alle $\lambda \in \mathbb{R}$ geldt:

$$\mathbb{E}\big((\lambda - X)^2\big) = \lambda^2 - 2\lambda\mathbb{E}(X) + \mathbb{E}(X^2) = \big((\lambda - \mathbb{E}(X))^2\big) + \mathrm{Var}\,(X) \ .$$

Bewijs van het Lemma. Zij $f : \mathcal{W}_Z \to \mathbb{R}$ zó dat $\mathbb{E}(f(Z)^2) < \infty$. Dan geldt:

$$\begin{aligned}
\mathbb{E}\big((f(Z) - X)^2\big) &= \sum_{z \in \mathcal{W}_Z} \mathbb{P}[Z = z] \cdot \mathbb{E}\big((f(z) - X)^2 | Z = z\big) \\
&= \sum_{z \in \mathcal{W}_Z} \mathbb{P}[Z = z] \cdot \big(f(z)^2 - 2f(z)\mathbb{E}(X|Z = z) + \mathbb{E}(X^2|Z = z)\big) \\
&= \sum_{z \in \mathcal{W}_Z} \mathbb{P}[Z = z] \cdot \big((f(z) - \mathbb{E}(X|Z = z))^2 + \mathbb{E}(X^2|Z = z) - \mathbb{E}(X|Z = z)^2\big) \\
&= \mathbb{E}\big((f(z) - \mathbb{E}(X|Z))^2\big) + \mathbb{E}\big(\mathbb{E}(X^2|Z) - \mathbb{E}(X|Z)^2\big) \ .
\end{aligned}$$

De tweede term hangt niet van f af, en de eerste is minimaal (namelijk 0) voor $f(z) = \mathbb{E}(X|Z = z)$, dus $f(Z) = \mathbb{E}(X|Z)$. □

Voor continu verdeeld stochasten, waar we in het volgende voornamelijk in geïnteresseerd zullen zijn, gaat definitie (1) van $\mathbb{E}(X|Z)$ niet meer op, omdat een gebeurtenis als $[Z = z]$ kans 0 heeft, zodat je er niet op kunt conditioneren. De karakterisering als gegeven in het Lemma blijft echter zinvol, en we zullen deze daarom gaan gebruiken als *definitie* van $\mathbb{E}(X|Z)$. Het zal handig blijken te zijn, meer dan één stochast te gebruiken als voorwaarde.

Definitie. Laten X en $Z_1, Z_2, \ldots, Z_n$ stochasten zijn met eindige variantie. Onder de *voorwaardelijke verwachting* $\mathbb{E}(X|Z_1, \ldots, Z_n)$ van X, gegeven $Z_1, Z_2, \cdots, Z_n$ verstaan we $f(Z_1, Z_2, \cdots, Z_n)$, waarbij (de meetbare functie) $f : \mathbb{R}^n \to \mathbb{R}$ zó gekozen is dat

$$\mathbb{E}\Big(\big(X - f(Z_1, Z_2, \cdots, Z_n)\big)^2\Big)$$

minimaal is.

3. Gaussische stochastische variabelen

Het Kalman-filter berekent voorwaardelijke verwachtingen in het heel speciale, maar belangrijke geval van normaal verdeelde stochasten. Het zal blijken dat bovenstaande definitie een mooie meetkundige interpretatie krijgt.

3.1. Een concreet model

Zij $N \in \mathbb{N}$. We maken een kansruimte $(\Omega_N, \Sigma_N, \mathbb{P}_N)$ als volgt. Voor Ω_N nemen we de N-dimensionale ruimte $\mathbb{R}^N$, en voor Σ_N de standaard-sigma-algebra die daarbij hoort. (We maken ons nu niet druk over de maattheorie van $\mathbb{R}^N$, en doen alsof we precies weten wat meetbare verzamelingen zijn. Voor de kenners: we nemen de Borel-verzamelingen. Voor anderen: we beschouwen alleen verzamelingen waar we over kunnen integreren.) Voor $\mathbb{P} = \mathbb{P}_N$ kiezen we die kansmaat op $\mathbb{R}^N$ die de coordinaten tot onafhankelijke normaal-verdeelde stochasten maakt: voor $A \subset \mathbb{R}^N$:

$$\mathbb{P}(A) := \frac{1}{(2\pi)^{n/2}} \int_A e^{-\frac{1}{2}(x_1^2+x_2^2+\ldots+x_N^2)} dx_1 dx_2 \cdots dx_N \ .$$

De stochastische variabelen die we zullen beschouwen zijn de *lineaire* functies $\Omega \to \mathbb{R}$, oftewel vectoren v in de *duale* ruimte Ω' van Ω:

$$X_v : \Omega \to \mathbb{R} : x \mapsto \langle v, x \rangle \ .$$

Deze keuze van stochasten heeft de volgende consequenties:

1. De stochast X_v is normaal verdeeld, met verwachting 0 en variantie $\|v\|^2$.
2. Als $v \perp w$, dan zijn X_v en X_w onafhankelijk: $X_v \perp\!\!\!\perp X_w$.

Deze twee eigenschappen vatten we samen met: De stochasten X_v, $v \in \Omega'$ zijn *gezamenlijk Gaussisch.*

Verdere gevolgen van onze constructie zijn:

3. De voorwaardelijke verwachting van X_v, gegeven X_z is X_{Pv}, waarbij P de orthogonale projectie is op de lijn door z.

3.2. Voorbeeld

Zij X een stochastische variabele met gemiddelde 0 en variantie a^2. Stel dat we X niet direct kunnen observeren, maar alleen met een normaal verdeelde waarnemingsfout W, die onafhankelijk is van X en die gemiddelde 0 en variantie m^2 heeft. We nemen dus de volgende stochast waar:

$$Z := X + W \ .$$

Wat is de beste schatting $\widehat{X}$ van X, gebaseerd op onze waarneming van Z?

Bewering. De beste schatting van X, gegeven Z is (zie figuur)

$$\widehat{X} = \frac{\mathbb{E}(XZ)}{\mathbb{E}(Z^2)} Z = \frac{a^2}{a^2 + m^2} Z \ .$$

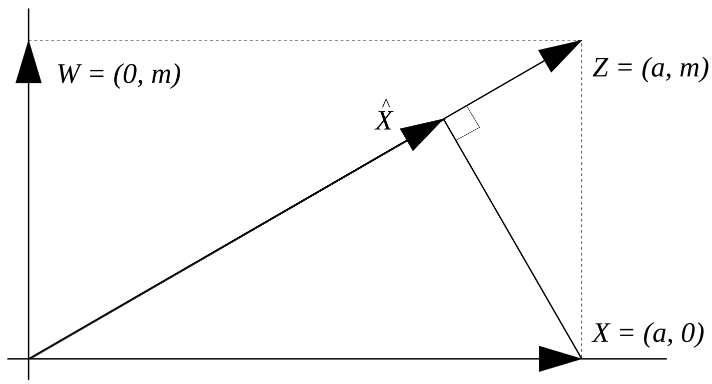
Bewijs. We modelleren deze situatie met de kansruimte $(\Omega_N, \Sigma_N, \mathbb{P}_N)$. Zij $N = 2$, $X = X_{(a,0)}$, $W = X_{(o,m)}$, zodat $Z = X_{(a,m)}$. De projectie van $(a, 0)$ op de lijn door 0 en (a, m) laat zich in het plaatje gemakkelijk aflezen:

$$p = \frac{a^2}{a^2 + m^2}(a, m) \ .$$

En dus is

$$\widehat{X} := \mathbb{E}(X|Z) = X_p = \frac{a^2}{a^2 + m^2} Z \ .$$

□



Je zou de factor $a^2/(a^2 + m^2)$ kunnen interpreteren als de 'geloofwaardigheid' van de waarneming Z. Deze brengt tot uitdrukking brengt in hoeverre wij de afwijking van Z van de a priori schatting 0 toeschrijven aan het voorwerp zelf, en in hoeverre we ze afdoen als ruis.

4. Een Kalman-filter

Neem aan dat de positie van een vliegend voorwerp ten tijde t gegeven wordt door

$$X + tV \ ,$$

waarbij X en V onafhankelijke, normaal verdeelde stochasten zijn, zeg met varianties respectievelijk α^2 en β^2, en met verwachting 0. Het idee is, dat α en β heel groot zijn, zodat wij van plaats en snelheid van het voorwerp op tijdstip 0 niet veel weten. De aanname $\mathbb{E}(X) = \mathbb{E}(V) = 0$ is hier enkel voor de eenvoud gemaakt.
Nu gaan we waarnemingen doen op tijdstippen $t_1, t_2, \cdots, t_n$; op tijdstip t_j meten we

$$Z_j := X + t_j V + R_j \ .$$

De *ruis* R_j is een rij onafhankelijke normaal verdeelde stochasten met verwachting 0 en variantie ε^2, onafhankelijk van X en V.
We modelleren dit probleem met behulp van Paragraaf 3.1.

Laat $N = n + 2$ en identificeer de volgende vectoren in Ω'_N met de stochasten uit het probleem.

$$\begin{aligned}
X &= \quad (\alpha, 0, 0, 0, 0, \ldots, 0) \\
V &= \quad (0, \beta, 0, 0, 0, \ldots, 0) \\
R_1 &= \quad (0, 0, \varepsilon, 0, 0, \ldots, 0) \\
R_2 &= \quad (0, 0, 0, \varepsilon, 0, \ldots, 0) \\
&\vdots \\
Z_1 &= \quad (\alpha, t_1\beta, \varepsilon, 0, 0, \ldots, 0) \\
Z_2 &= \quad (\alpha, t_1\beta, 0, \varepsilon, 0, \ldots, 0) \\
&\vdots
\end{aligned}$$

Zij nu $\mathscr{Z}_j$ de lineaire ruimte, opgespannen door $Z_1, Z_2, \ldots, Z_j$. Zij P_j de orthogonale projectie van Ω' op $\mathscr{Z}_j$. Dan is de beste schatting $\widehat{X}_j := \mathbb{E}(X|Z_1, Z_2, \cdots, Z_j)$ de projectie van $X = X_{(\alpha,0,0,\ldots,0)}$ op $\mathscr{Z}_j$: $\widehat{X}_j = P_j(X)$, en evenzo is $\widehat{V}_j = P_j(V)$.
Op tijdstip 0 hebben we nog niets gezien, en schatten we X en V met hun verwachtingswaarden:

$$\widehat{X}_0 = 0 \;, \quad \widehat{V}_0 = 0 \;.$$

Op tijdstip t_1 zien we Z_1, en we *updaten* onze kennis over X en V:

$$\widehat{X}_1 = c_1 Z_1, \quad \text{met } c_1 \text{ zó dat} \quad X - \widehat{X}_1 \perp Z_1 \;.$$

Dus moet $\langle X - c_1 Z_1, Z_1 \rangle = 0$, zodat $\langle X, Z_1 \rangle = c_1 \langle Z_1, Z_1 \rangle$, dat wil zeggen: $\alpha^2 = c_1(\alpha^2 + \beta^2 t_1^2 + \varepsilon^2)$. We vinden:

$$\widehat{X}_1 = \frac{\alpha^2}{\alpha^2 + \beta^2 t_1^2 + \varepsilon^2} Z_1 \;,$$

en analoog:

$$\widehat{V}_1 = \frac{\beta^2 t_1^2}{\alpha^2 + \beta^2 t_1^2 + \varepsilon^2} Z_1 \;.$$

De volgende stap, na de waarneming van Z_2, is ingewikkelder dat de eerste, omdat Z_1 en Z_2 niet onafhankelijk (orthogonaal) zijn. Het is dan voordelig, in stap $j \geq 2$ eerst het *innovatieve* of verrassende deel N_j uit de nieuwe waarneming Z_j te isoleren:

$$N_j := Z_j - P_{j-1}(Z_j) \;.$$

Deze *innovaties* $N_1, N_2, \cdots, N_n$ zijn onderling onafhankelijk, en, gezien als vecoren in Ω', orthogonaal. Hieruit volgt dat

$$\widehat{X}_j = \widehat{X}_{j-1} + \frac{\mathbb{E}(XN_j)}{\mathbb{E}(N_j^2)} N_j \quad \text{en} \quad \widehat{V}_j = \widehat{V}_{j-1} + \frac{\mathbb{E}(VN_j)}{\mathbb{E}(N_j^2)} N_j \;. \tag{2}$$

Op basis hiervan komen we tot het volgende resultaat

Stelling 2. *De beste schattingen van X en V, gegeven de eerste j waarnemingen worden gegeven door de volgende recursieve vergelijkingen.*

$$\begin{aligned} \widehat{X}_j &= \widehat{X}_{j-1} + \frac{\alpha_{j-1}^2 + t_j\gamma_{j-1}}{\alpha_{j-1}^2 + t_j^2\beta_{j-1}^2 + 2t_j\gamma_{j-1} + \varepsilon^2}\left(Z_j - \left(\widehat{X}_{j-1} + t_j\widehat{V}_{j-1}\right)\right) ; \\ \widehat{V}_j &= \widehat{V}_{j-1} + \frac{t_j\beta_{j-1}^2 + \gamma_{j-1}}{\alpha_{j-1}^2 + t_j^2\beta_{j-1}^2 + 2t_j\gamma_{j-1} + \varepsilon^2}\left(Z_j - \left(\widehat{X}_{j-1} + t_j\widehat{V}_{j-1}\right)\right) , \end{aligned} \tag{3}$$

waarin de parameters α_j, β_j en γ_j worden berekend met de recursie

$$\begin{aligned} \alpha_0 = \alpha ; &\quad \alpha_j^2 = \alpha_{j-1}^2 - \frac{(\alpha_{j-1}^2 + t_j\gamma_{j-1})^2}{\alpha_{j-1}^2 + t_j^2\beta_{j-1}^2 + 2t_j\gamma_{j-1} + \varepsilon^2} ; \\ \beta_0 = \beta ; &\quad \beta_j^2 = \beta_{j-1}^2 - \frac{(t_j\beta_{j-1}^2 + \gamma_{j-1})^2}{\alpha_{j-1}^2 + t_j^2\beta_{j-1}^2 + 2t_j\gamma_{j-1} + \varepsilon^2} ; \\ \gamma_0 = 0 ; &\quad \gamma_j = \gamma_{j-1} - \frac{(\alpha_{j-1}^2 + t_j\gamma_{j-1})(t_j\beta_{j-1}^2 + \gamma_{j-1})}{\alpha_{j-1}^2 + t_j^2\beta_{j-1}^2 + 2t_j\gamma_{j-1} + \varepsilon^2} . \end{aligned}$$

Bewijs. Definieer

$$\alpha_j^2 := \|X - \widehat{X}_j\|^2 ; \quad \beta_j^2 := \|V - \widehat{V}_j\|^2 ; \quad \gamma_j := \langle X - \widehat{X}_j, V - \widehat{V}_j \rangle .$$

Dan geldt:

$$\begin{aligned} \mathbb{E}(XN_j) &= \langle X, (\mathbf{1} - P_j)Z_j \rangle \\ &= \langle X - \widehat{X}_{j-1}, (X - \widehat{X}_{j-1}) + t_j(V - \widehat{V}_{j-1}) + R_j \rangle \\ &= \|X - \widehat{X}_{j-1}\|^2 + t_j\langle X - \widehat{X}_{j-1}, V - \widehat{V}_{j-1} \rangle \\ &= \alpha_{j-1}^2 + t_j\gamma_{j-1} . \\ \mathbb{E}(N_j^2) &= \|(X - \widehat{X}_{j-1}) + t_j(V - \widehat{V}_{j-1}) + R_j\|^2 \\ &= \alpha_{j-1}^2 + t_j^2\beta_{j-1}^2 + 2t_j\gamma_j + \varepsilon^2 . \end{aligned}$$

Met (2) volgt nu (3). De recursies voor α_j, β_j en γ_j worden nu gevonden met

$$\begin{aligned} \alpha_{j-1}^2 - \alpha_j^2 &= \langle X, (P_j - P_{j-1})X \rangle = \frac{\mathbb{E}(XN_j)^2}{\mathbb{E}(N_j^2)} ; \\ \beta_{j-1}^2 - \beta_j^2 &= \langle V, (P_j - P_{j-1})V \rangle = \frac{\mathbb{E}(VN_j)^2}{\mathbb{E}(N_j^2)} ; \\ \gamma_{j-1} - \gamma_j &= \langle X, (P_j - P_{j-1})V \rangle = \frac{\mathbb{E}(XN_j)\mathbb{E}(VN_j)}{\mathbb{E}(N_j^2)} . \end{aligned}$$

□