

Kalman Filter für Dummies

“Je älter wir werden, desto mehr steigt unser Anspruch, aus dem Meer der Belanglosigkeiten die wichtigen Dinge herauszufischen. Diesen Vorgang nennen die Techniker Filtern.”

—Hans Weigl

Was?

Das Thema Kalmanfilter ist nicht so einfach zu erklären, wenn man allzu mathematisch denkt. Dann wird es in der Tat beliebig kompliziert. Natürlich geht es auch nicht ohne Mathematik. Witzigerweise leben wir tagtäglich mit Kalmanfiltern, ohne sie zu bemerken. Wir vertrauen darauf sogar beinahe blind. Wenn wir nämlich das Navigationssystem im Auto laufen haben, uns nach links oder rechts leiten lassen und den Ansagen desselben brav folgen. Dann nämlich arbeitet in dem Gerät tatsächlich ein Kalmanfilter, der sich durch die Livedaten der Satelliten kämpft und die optimale Position errechnet. Nicht zu fassen, aber wahr! Erfahrungsgemäss läuft das alles gar nicht mal so schlecht.

Um es kurz einmal anzudeuten: das Ganze ist eine Menge von Gleichungen mit ein paar indizierten Variablen, die via Matrizen sukzessive weitergerechnet werden, indem ausserdem Messdaten von aussen dazugespielt werden. Ich weiss, das war jetzt reichlich dunkle Prosa. Aber wir wollen Licht in das Dunkel bringen. Die Lernkurve ist beim Kalmanfilter anfangs recht flach, das wissen wir aus eigener Erfahrung. Aber nach und nach fällt es einem wie Schuppen von den Augen.

Und damit wünsche ich viel Vergnügen! Übrigens schreiben wir ab jetzt kurz KF anstelle von Kalmanfilter.

Problemstellung

Angenommen man hat ein Signal. Irgendein Signal, das von aussen kommt und sich als Reihe von Messwerten auftragen lässt. Beispielsweise Radarsignale oder sukzessive Farbwerte beim Scannen eines Bildes. Eines ist immer der Fall: die Werte sind ungenau, d.h. verrauscht. Rauschen ist ein Bestandteil unserer Umwelt. Jedes Gespräch findet vor einer Kulisse aus Geräuschen statt, aber der Zuhörer überhört das Unwesentliche und nimmt nur das wahr, was ihm Information bedeutet. Was also kann man tun, um das Rauschen loszuwerden? In der Charttechnik fällt den Leuten meist nur der gleitende Durchschnitt ein. Das ist natürlich das Ungeschickteste, was man tun kann. Die meisten Probleme der Technik lassen sich nicht so einfach aus der Welt schaffen. Hier muss etwas Ausgeklügeltes her, etwas das richtig stark darin ist, Rauschen zu eliminieren und Kurs zu halten.

Highlights

- Ein Filter ist genau genommen ein Estimator.
- Es heisst im übrigen **das** Filter und nicht **der** Filter. Der Filter wäre beispielsweise der Kaffeefilter.
- Das KF ist reichlich innovativ und wirksamer als man zuerst glauben mag
- Das KF ist eine der wichtigsten Entdeckungen im letzten Jahrhundert
- Mit ein wenig Mathematik kann man bereits erste KFs bauen, aber um ihn komplett zu beherrschen, müsste man ein John von Neumann sein
- Es eignet sich wirklich gut zum implementieren als Computer Algorithmus
- Es ist rekursiv, d.h. der Output wird im Folgeschritt zum Input

Q&A

“Kann man das KF auf alle digitalen Signale losslassen?”

Wenn man imstande ist, ein schlaues Modell zum Problem zu bauen, das ein geschlossenes KF ergibt, so sollte das immer funktionieren.

“Kann man das KF für Image Processing benutzen? ”

Ja natürlich.

Wer?

Anlässlich einer würdigen Zeremonie im “East Room” des Weissen Hauses am 7. Oktober 2009, verlieh U.S. Präsident Barack Obama dem Ingenieur Rudolf E. Kalman von der ETH Zürich die “2008 National Medal of Science”. Während des Festaktes wurden die Auszeichnungen “2008 National Medal of Science” sowie die “National Medal of Technology and Innovation” an bedeutende Wissenschaftler verliehen.



Figure 1.1. Obama ehrt in 2009 R. Kalman mit der National Medal Of Science

Rudolf Emil Kalman wurde in Budapest, Ungarn, geboren und erhielt seinen Hochschulabschluss im Jahre 1953. 1954 erhielt er seinen Master-Titel am MIT (Massachusetts Institute of Technology) auf dem Gebiet der Elektrotechnik. Der Doktor-Titel folgte dann 1957 an der Columbia University. Kalman ist also ausgebildeter Elektroingenieur und berühmt für seine Erfindung des Kalman Filters, eine mathematische Technik, die weitverbreitet ist in Messsystemen und der Luftfahrtelektronik. Diese Technik erlaubt es, ein brauchbares Signal aus einer Reihe von unvollständigen und verrauschten Messwerten zu bestimmen. Kalmans Idee wurde zu Beginn mit sehr grosser Skepsis betrachtet, so dass er gezwungen war, seine Ergebnisse zuerst in einer Maschinenbau-Zeitschrift anstelle einer mathematischen zu veröffentlichen. Seine Idee war jedoch erst erfolgreich, als er 1967 Stanley F. Schmidt vom NASA Research Center besuchte

und ihm die Filtertechnik präsentierte. Das hatte dann zum Einsatz des Kalman Filters während des Apollo-Programms geführt!

Wie?

Quick & Dirty

Natürlich lässt sich die volle Bedeutung des KF ohne Mathematik nicht erfassen. Beginnt aber mit zuviel Mathematik steht man im Wald. Wir skizzieren hier die wesentlichen Gleichungen und deuten diese. Das ist für ein Anfangsverständnis hilfreich.

$$\hat{x}_k = G_k v_k + (1 - G_k) \hat{x}_{k-1}$$

\hat{x}_k ist der Schätzwert für heute, und \hat{x}_{k-1} ist der Schätzer von gestern

G_k ist das Kalman Gain

v_k ist der Messwert für heute

Null

Die k's im Subskript bedeuten Zustände, die einander nachfolgen. Wir können uns hier einmal Zustände zu den diskreten Zeiten $k=1, k=2, \dots$ vorstellen. Konkreter gesprochen wären die Zustände einer Aktie, z.B. ihr Kurs v_k zur Eröffnung an jedem aufeinanderfolgenden Tag k .

Unser Ziel ist es, den Schätzwert \hat{x}_k für das Signal v_k zu finden und zwar für jeden aufeinanderfolgenden Tag (oder eine andere Zählheit). Man kann nun einwenden, warum den Schätzwert für den Kurs v_k berechnen, wenn man ihn schon hat. v_k ist zwar ein echter Kurswert, aber er ist verrauscht. Es gibt ja noch den High, den Low und den Closekurs uvm. und welcher davon ist nun der 'beste' Kurs im Tagesverlauf? Was aber verstehen wir nun unter 'der Beste'? Im Sinne der Navigation der Stabilste, im Sinne des Börsenhandels den, der z.B. die beste Rendite beim Ordern verspricht. Wir sind uns also nicht sicher bezüglich des Messwertes v_k . G_k ist das sogenannte 'Kalman Gain'. Das ist der Schlüssel, um weiterzukommen. \hat{x}_{k-1} ist übrigens der Schätzer von gestern. Der wird nun zum Input. Und so geht es Schritt für Schritt weiter.

Die grosse Unbekannte ist G_k . Wir haben die Messungen v_k und wir haben alle vergangenen Schätzer. Damit kann man das Kalman Gain für jeden Schritt neu berechnen. Das ist nicht ganz einfach und wird unten (ein bisschen) gezeigt. Wäre das Kalman Gain immer 0.5, dann hätten wir in der Tat die übliche Durchschnittsberechnung. Aber so dumm ist das KF natürlich nicht. Es gibt definitiv smartere G_k .

Was tut das KF nun? Wir fassen einmal kurz zusammen:

Das KF findet den optimalsten Gewichtungsfaktor, um Durchschnitte zu bilden für jeden aufeinanderfolgenden Zustand, den das beobachtete Objekt annimmt. Und dazu erinnert es sich ein bisschen an die vergangenen Signalwerte.

Verblüffend, nicht wahr!

Step by step Guide

Step 1- Ein Modell aufstellen

Das ist der wichtigste Punkt überhaupt. Man muss sich sicher sein, dass das Modell dem Problem, mit dem man sich herumschlägt, in etwa entspricht. Dann nämlich hat man die Rahmenbedingungen für das KF geschaffen, in denen es sich bewegt. Unter Modell verstehen wir z.B. ein physikalisches Modell, wenn wir ein Fahrzeug steuern wollen oder ein rein mathematisches Modell der Ökonomie, wenn wir ökonomische Zeitreihen beobachten. Also haben wir einen Modellwert x_k , der sich quasi hinter dem Schätzer \hat{x}_k verbirgt und als Stellvertreter für den besten Messwert angesehen werden kann, den wir jedoch nur verrauscht erhalten (das wären unten die z_k). Ein Modell könnte nun sein

$$x_k = Ax_{k-1} + Bu_k + n_{k-1}$$

A und B sind Matrizen, u_k ist ein Steuersignal und n_{k-1} ist der Rauschanteil des Prozesses

und

$$z_k = Hx_k + m_{k-1}$$

m_{k-1} ist ein Rauschterm und z_k ist der verrauschte, beste Modellwert

Also ist jeder Modellwert x_k (der im Modell stellvertretend für den besten Messwert, der eben nicht v_k ist, gedacht werden kann) eine lineare stochastische Gleichung. Jedes x_k ist eine Linearkombination aus seinem vergangenen Wert plus einem Kontrollwert inklusive einem Rauschanteil. Das Kontrollsignal könnte nun ein korreliertes Signal sein. Man denke z.B. an dem DAX, der fast immer den Vorgaben aus Amerika folgt. Man kann der Einfachheit halber den Kontrollanteil aber auch erst einmal weglassen.

Die zweite Gleichung bestimmt den Modellwert x_k mit Verrauschung. Die Rauschanteile werden oft als gaussch definiert. Ausserdem sollen beide Rauschterme m_{k-1} und n_{k-1} (statistisch) unabhängig voneinander sein. A , B und H sind in der Regel Matrizen. Aber das KF lässt sich oft so einfach formulieren, dass es nur numerische Werte sind. A , B und H können mit der Zeit variieren, aber das lassen wir hier erst einmal weg.

Das Modell oben ist übrigens so allgemein, dass die meisten Probleme damit beschrieben werden können. Nun müssen wir noch den Mittelwert und die Standardabweichung der Rauschanteile bestimmen. Im wahren Leben ist zwar nichts gaussch, aber in Approximation eben doch. Das KF ist so stabil, dass sogar ganz schlechte Ansätze für die Rauschterme wenig Einfluss haben auf die Qualität der Schätzer, wenn das KF erst einmal angelaufen ist. Am Anfang zeigen sich ein paar Ausschläge, die sich aber schnell beruhigen. Trotzdem gilt: **Je besser der Rauschansatz ist, desto besser sind die Schätzer.**

Step 2 - Den Prozess starten

Hier nun starten wir die Iterationen des KF. Vor allem brauchen wir gescheite initiale Werte und Parameter. Es wechseln sich Prädiktion und Korrektur mit einer neuen Messung ab. Danach folgt die nächste Prognose, die wiederum mit dem folgenden Messwert korrigiert wird. Undso weiter undso weiter. Der Unterschied liegt in der vorhandenen Information: die Prognose hat weniger $(k - 1)$ Messwerte zur Verfügung als die nachfolgende Korrektur. Die hat nämlich k Messwerte im Gedächtnis. Diesen wesentlichen Unterschied drücken wir in der Notation- $\hat{x}_{k,k-1}$ und $\hat{x}_{k,k}$ aus.

Prädiktion

$$\hat{x}_{k,k-1} = A \hat{x}_{k-1,k-1} + B u_k$$

$\hat{x}_{k,k-1}$ ist der Schätzer, der auf Messinformationen bis gestern $(k - 1)$ beruht. Deswegen sprechen wir von Prädiktion.

Korrektur

Das Update der Prädiktion mit dem neuen verrauschten Messwert z_k ergibt den Schätzer $\hat{x}_{k,k}$ der auf den Informationen zum Zeitpunkt k beruht, also mehr weiss als sein Vorgänger $\hat{x}_{k,k-1}$.

$$\hat{x}_{k,k} = \hat{x}_{k,k-1} + G_k (z_k - H \hat{x}_{k,k-1})$$

Die Matrizen A , B und H sind bekannt. $\hat{x}_{k,k}$ modelliert man in jedem Schritt aus den Matrizen H und dem verrauschten Messwert z_k zusammen mit dem Kalman Gain. G_k wird aus H und noch zusätzlichen Matrizen, die die Rauschanteile zum Ausdruck bringen, bestimmt. Und dies ebenfalls iterativ, also in jedem Durchlauf des KF neu. Die schmerzliche Mathematik hierzu vertagen wir auf später. Anstelle der Matrizen kann man einfach Zahlen einsetzen und so sehen, welche rudimentäre Form das KF annehmen kann (und dabei immer noch effektiv bleibt). Um zu starten, brauchen wir ein $\hat{x}_{0,0}$. Dazu kann man irgendeinen Wert nehmen. Am besten (im Fall einer Aktie) den ersten Kurs der Zeitreihe oder die erste Kursdifferenz, wenn man das Momentum (sprich die erste Ableitung) in das Modell integriert hat.

Step 3 - iterieren, iterieren, iterieren ...

Die beiden Gleichungen aus Schritt zwei werden nun immer wieder durchlaufen. So oft eben wie man Messwerte zur Verfügung hat. Die Zeitreihe der $\hat{x}_{k,k}$ ist dann die Zeitreihe der optimalen Werte des Ausgangsobjekts. Bedenken Sie, dass der vergangene Schätzer immer der Input des nächsten ist. $\hat{x}_{k,k-1}$ ist einfach die gröbere Variante von $\hat{x}_{k,k}$, da weniger Information zur Verfügung steht. Um noch ein schönes Wort einzuführen: die $\hat{x}_{k,k-1}$ sind die a priori Schätzer, die anderen folgen aposteriorisch. Das macht Sinn!