The Kalman Filter

A. V. Balakrishnan

The 'Kalman Filter' is best described as a soft-ware data-processor designed to filter signal from noise. It has been hailed as the single specific advance in the Systems area with significant impact on the battle front, according to at least one purported Air Force survey. Beginning with its application to satellite orbit tracking, it is now in routine use in all Navigation (position location) Systems: Inertial Navigation, Satellite Navigation, Radio Navigation as well as Radar and Sonar Tracking Systems. In military applications it is an integral part of Guidance Systems (where essentially the missile is made to follow a desired trajectory). In the near future it may well be employed in all automobile ignition systems.

For the mathematician it is just a sub-topic in statistics—stochastic processes — an area in which the revered mathematical names are Wiener and Kolmogorov. And yet it is remarkable that this particular theory did not originate from any traditionally trained Probabilist, etc., but rather from a 'system scientist' R. E. Kalman. [It is curious that there is a similar parallel in Information Theory — that after the pioneering work of Claude Shannon, despite the efforts of many bona-fide statisticians, by and large, the only significant (to communication engineering) result was that of A. Viterbi, a communication engineer who developed the decoding scheme bearing his name for convolutional codes.]

As part of the theory of stochastic processes, the Kalman Filter, while of immediate practical utility, can be described in 'axiomatic' mathematical language, and indeed rather sophisticated mathematics at that. In this respect, it is unlike the traditional view among mathematicians of "application" as mathematical physics, boundary value problems and associated numerical techniques. The general image that many mathematicians have of "Applied Mathematics" is basically of 'sticking numbers' into 'formulae'. The Kalman Filter is the best illustration of how wrong this view can be. It does not "solve" any equation and has nothing to do with physics. It does come up with a "formula" or "recipe" for processing data but 'application' does not mean merely putting numbers into formulae — indeed any real discussion of a Kalman Filtering application requires no less than a 'case study' full of the detailed and specialised jargon of the application area involved.

Central to the success of the Kalman Filter are two ideas: one is the exploitation of the 'dynamic system' representation of the signal and the other is the 'white-noise' representation of the noise. The early notion of a filter was simply one that 'cuts off' all frequencies beyond some value that did not distort the signal. Then came the theories associated with the names of Wiener and Kolmogorov. These theories were in a sense too sophisticated — the models as well as the results were too general to be of value in most applications. Indeed it is an oft repeated cliche in communication engineering that no Wiener filter was ever implemented.

It is necessary now to describe, however briefly, the mathematical theory involved so that we can be more precise. The Kalman Filter theory can be developed for 'continuous-time' or 'discrete-time' ('time-series') models. Since all Kalman filtering currently involves digital computation, the latter may be thought to be more appropriate but we shall see that this is not necessarily so. Since this note is addressed to mathematicians, we shall be concerned primarily with the basic assumptions and results rather than with the details of the derivation. We begin with the discrete-time theory in its most useful — rather than the most general — form.

Let $\{y_n\}$ denote the 'observed data' (sensor output, for example) sequence which is assumed to have the structure:

$$y_n = S_n + N_n$$

where $\{N_n\}$ is 'white noise': that is to say, N_n is a sequence of independent zero mean and known variance (and hence taken to be unit for simplicity) Gaussian variables, and $\{S_n\}$ is the "information-bearing" signal. Here $\{N_n\}$ represents the unavoidable 'instrument' error that remains even after all known sources of 'systematic' errors have been corrected for. An equally important assumption is that the sequence $\{S_n\}$ has the 'dynamic' representation

$$S_n = Cx_n$$

$$x_{n+1} = \Phi x_n + F\zeta_n$$

where C, Φ, F are known matrices and $\{\zeta_n\}$ is white noise again, and (for simplicity) statistically independent of the

noise sequence $\{N_n\}$. The Kalman Filter is designed to provide the 'least-squares' estimate of S_n based on data observed up to n: that is, to calculate

$$\hat{S}_n = E[S_n]y_m, \quad m \leqslant n]$$

The remarkable consequence of the assumed model (in contrast to the more general models of Wiener) is that we may express this estimate in a recursive way ('updating'):

$$\hat{S}_n = C\hat{x}_n$$

$$\hat{x}_{n+1} = \Phi\hat{x}_n + K_{n+1}[y_{n+1} - C\Phi\hat{x}_n]$$

Here $\hat{x}_n$ is the current estimate and the second term represents the "updating" based on the difference from the 'new piece of data' y_{n+1}. The advantage of the recursive estimate is that all the past data need not be stored. The sequence $\{K_n\}$ does not depend on the data and can be calculated "off-line", and moreover (under usual conditions) its asymptotic (large n) value is a calculable constant. Above all is the simplicity of structure of the filter.

The continuous-time version is similar but does involve a mathematical difficulty concerning the representation of the observation noise. Thus let

$$y(t)$$

represent the observed data and assume that

$$y(t) = S(t) + N(t)$$

where $S(t)$ is the signal characterised by the model:

$$S(t) = Cx(t)$$

$$\dot{x}(t) = Ax(t) + Fn(t)$$

where n(t) is "white noise' of unit spectral density (matrix), A and F are given matrices. The observation noise N(t) is also taken to be white independent of the process n(t) and also of unit spectral density. Note that unlike the discrete case, the 'white-noise' processes are mathematically awkward because they have infinite variance. It is currently in vogue to use the 'integrated version':

$$Y(t) = \int_0^t y(s)ds = \int_0^t S(\sigma)d\sigma + W(t)$$

where W(t) is a Wiener process, as being "more rigorous". This is nonsense because no physical instrument can output a Wiener process. The noise process is simply an idealization of the fact that the spectral bandwidth must be much larger than that of the signal and 'flat' over this range if the instrument is not to 'distort' the signal. More on this below. In any event, defining (with suitable interpretation):

$$\hat{S}(t) = E[S(t)|y(\sigma), \sigma \leqslant t]$$

$$\hat{x}(t) = E[x(t)|y(\sigma), \sigma \leqslant t]$$

we have

$$\hat{S}(t) = C\hat{x}(t)$$

$$\hat{x}(t) = A\hat{x}(t) + P(t)C^*[y(t) - C\hat{x}(t)]$$

$$= (A - P(t)C^*C)\hat{x}(t) + P(t)C^*y(t)$$

the second form being particularly important in the asymptotic version $(t \to \infty)$ when P(t) is a constant.

There is a dilemma faced by every knowledgeable designer as to which model to use — discrete or continuous. If we assume the discrete model we must sample the data slowly enough to insure that the noise samples are indeed independent, but only the noise r.m.s. value need be known. On the other hand if we assume the continuous model it implies that we may sample that data at as high a rate as we wish but now we must know the spectral density. Here is one illustration of the point made earlier — 'application' is no longer just putting numbers into formulae. Rather difficult and often subtle decisions have to be made concerning the model to be used! Then again there is often the equally difficult question: how does one make sure one is right? The usual way out is to use a "computer simulation" which actually proves little, being a tautologous procedure. It is interesting in this connection to mention that just as the mathematicians misconstrue the Kalman Filter because it uses mathematics, in the same way the engineering users (most of whom have only a dim understanding of the theory) tend to look upon the Kalman Filter as a magic black box — and think that 'twiddling the dials' enough they can make it "work" — in spite of the theoreticians!

Of course in all the applications the 'linearity' of the equations is only an approximation and the instrument noise is not really white (or even Gaussian). The filter is "robust" enough to be good in spite of these limitations within reasonable bounds. What these bounds are — the judgment between too complex (and intractable) a model and too simple (and hence inadequate) is what the

designer's ingenuity must settle. No mathematician who has not had his "hands dirty" with a real, practical problem of this sort can appreciate what is involved – what "systems" work means – and herein appears again the difference from the usual "boundary-value problem numerical analysis" view of applications. This is of course not to say that significant – even great – contributions cannot come from non-practitioners.

The "brittleness" of the noise model is made more apparent in the push toward extending this theory to 'non-linear' dynamic models for the signal. Indeed the Wiener process models lead not only to contradictions but also to formulas which are meaningless – because the assumption that the 'integrated version' contains a Wiener process is simply the wrong extension of the model! However this cannot be appreciated by the "Non-Linear Filter" theorists who have never bothered to use any of their results on any 'real' data. Of course the fact that much of the current theory uses models which are not appropriate is not necessarily an argument against it!

References

The basic literature:

1. Kalman, R. E.: *A New Approach to Linear Filtering and Prediction Problems, Journal of Basic Engineering, Vol. 1, 1960, p. 597–622. (where it all began)*
2. Battin, R.H.: Astronautical Guidance, McGraw Hill, 1964. (Space Guidance Application)
3. Bucy, R. S. and Joseph, P. D.: *Filtering for Stochastic Processes with Applications to Guidance, Interscience N.Y., 1968. (Source Book)*
4. Meditch, J. S.: *Stochastic Optimal Linear Filtering and Control, McGraw Hill, 1969. (Discrete Case)*

For a detailed account of the Wiener Process Version of Linear and Non-Linear Filtering:

1. Liptser, R. S. and Shiryayev, A. N.: *Statistics of Random Processes, Vol. I, and II – Springer-Verlag 1977.*

Acknowledgment

System Science Department
University of California
Los Angeles, USA

New Undergraduate Texts

Basic Concepts of Algebraic Topology is designed for use as an introductory text at the undergraduate and beginning graduate levels. It emphasizes a geometric approach and may be considered a reaction against the high-powered axiomatic method of teaching algebraic topology. Concepts are illustrated with simple examples, and the proofs originally given by the discoverers of the important theorems are used when they are consistent with the introductory level of the course.

This innovative text contains the topics usually covered in a full years' course in finite mathematics or mathematics for liberal arts students. The exposition of the mathematical material is accompanied by an explanation of the computer language APL which then, in turn, is used for the treatment of applications and, in general, for a deeper understanding of the theory. The many problems accompanying the text lead the student to a more thorough understanding.

This book contains a sophomore level course in linear algebra and presents the basic, essential facts, as they will be needed to be built upon in further mathematics courses. Its goal is the derivation and proof of the Principal Axes Theorem for symmetric linear transformations.
There is a wealth of numerical examples which illustrate the more abstract parts of the theory. More than 200 exercises are included.
Prices are subject to change without notice