

KALMAN FILTERS AND ARMA MODELS

Aniello FEDULLO¹

Abstract. The Kalman filter is the celebrated algorithm giving a recursive solution of the prediction problem for time series. After a quite general formulation of the prediction problem, the contributions of its solution by the great mathematicians Kolmogorov and Wiener are shortly recalled and it is showed as Kalman filter furnishes the optimal predictor, in the sense of least squares, for processes which satisfy the linear models with a finite number of parameters, that are the ARMA models.

1. Introduction: Time Series

A *time series*, in our study, is considered like a finite part (*a sample*) of a single realization of a stochastic process. The fundamental problem of the analysis of the time series is the following: given a time series, infer, at least in part, the characteristics of the process. Remember that, in general, a stochastic process is characterized by the joint distributions of all the finite sub-families of its random variables. If there is no other information known on the process, or if no other hypothesis is made about it, the problem is unsolvable or perhaps ill-posed. Conversely, if we limit ourselves to particular families of processes, the above-mentioned statistical inference is possible. In particular this happens for *weak (second order) stationary, ergodic, invertible* and with a *Gaussian residual* processes (see Papoulis 1965). In the following, unless otherwise indicated, we will consider only such processes.

2. The prediction problem

The information on the process inferred from the time series allows to resolve problems particularly important for applications, such as those of prediction, filtration and control. In the present work we will concentrate our

¹ Dipartimento di Matematica e Informatica, Università di Salerno, Italia.

attention on the first of the three. The problem of prediction is to estimate a future value $X(t+m)$ ($m>0$) of a time series whose past values $X(t)$, $X(t-1)$, ..., $X(t-n)$ are known. The estimation $\hat{X}^{(t+m)}$ of $X(t+m)$ will be a suitable function of these latter values which minimizes

$$M(m) := E[(\hat{X}^{(t+m)} - X(t+m))^2].$$

It can be shown that in the absence of imposed constraints on the form of the aforementioned function, the preceding minimum problem has the solution given by the *conditional expectation*

$$E[X(t+m) | X(t), X(t-1), \dots, X(t-n)].$$

The calculation of the latter, in general, requires the knowledge of the joint distribution of the random variables involved, that is a very detailed knowledge of the process, one rather difficult to arrive at. Observe that if the joint distributions are *normal* then the conditional expectation is linear, for which the so called *predictor* is written

$$\hat{X}^{(t+m)} = a_0 X(t) + a_1 X(t-1) + \dots + a_n X(t-n)$$

where a_0, \dots, a_n are constants (dependent only on m and n) to determine. This is done in such a way to minimize M and in order to do so it suffices to know the *autocovariance*. If the process is Gaussian then the linear predictor (2) is *optimal*, in the sense of least squares. Otherwise the conditional expectation is not in general linear and, as said, we could not know how to calculate it easily. Nevertheless, in such a case, we can restrict to the aforementioned linear predictors and search among them for those which are optimal. The originators of the researches on linear predictors were Kolmogorov and Wiener in the early '40s. The publication of their works, however, was held up until 1949 because of military concerns (automatic pointing of anti-aircraft weapons and fire control). It became clear then that Kolmogorov and Wiener had resolved independently the same problem, using different techniques. They, as it was understood afterwards, related to different choices of the coordinates related up to the same geometric problem in Hilbert space. Furthermore, both the authors also studied the semi-infinite version of the problem ($n \rightarrow \infty$), for the major simplicity of the mathematical treatment. That, nevertheless, can be used as a good approximation of the real case where n is large but finite.

3. Recursive Algorithms

Kolmogorov's and Wiener's approaches to the problem of prediction has the advantage of furnishing an explicit expression of $X^{(t+m)}$, but a notable disadvantage is that it must be recalculated ex novo for different values, also consecutive, of m . This carries enormous complications in the computations, especially for real time applications. Such problems were overcome by the development of recursive algorithms. These calculate the estimation at time $n+1$ by means of a simple correction of that at time n , with a notable saving of memory and of computation time. The two most famous recursive approaches are the one by Box and Jenkins and the one by Kalman. Both conducted to the optimal linear predictor in the sense of least squares, and both are applied only to the processes which satisfy the linear models with a finite number of parameters.

4. Linear Models with a Finite Numbers of Parameters

a) ARMA models

Let $X(t)$ be a stochastic process which satisfies the properties indicated at the end of section 3 and which we assume, without loss of generality, has mean 0. Let $h(\omega)$ be the *spectral density* of $X(t)$; that is a non negative real function of ω which we will also assume to be a rational function of $\exp(i\omega)$. For the *spectral factorization theorem* there exist the polynomials $A \in C$ such that

$$h(\omega) = \text{const} \frac{A(e^{i\omega})A^*(e^{i\omega})}{C(e^{i\omega})C^*(e^{i\omega})}$$

where

$$A(z) := \sum_{i=0}^p a_i z^i, \quad A^*(z) := \sum_{i=0}^p a_i z^{p-i}; \quad a_0 := 1; \quad a_i \in \mathbb{R} \text{ for every } i$$

$$C(z) := \sum_{i=0}^q c_i z^i, \quad C^*(z) := \sum_{i=0}^q c_i z^{q-i}; \quad c_0 := 1; \quad c_i \in \mathbb{R} \text{ for every } i$$

We observe that $A(w) = 0$ if and only if $A(1/w) = 0$, by which, in virtue of the *fundamental theorem of algebra*, we can choose A and C in such a way that their zeros are all in modulus greater than 1. Then the process $X(t)$ can be written as

$$X(t) = \frac{A(\beta)}{C(\beta)} e(t)$$

where β is the *shift operator* defined as

$$\beta X(t) := X(t-1) \quad \text{for every } t$$

and $e(t)$ is a *white noise*, that is a noise made up of random variables with mean 0, variance σ^2 and pairwise uncorrelated. In fact it is easy to verify that the right hand of (4) has the spectrum (3). Multiplying both sides of (4) by C one has

$$C(\beta)X(t) = A(\beta)e(t)$$

called *autoregressive moving average* or ARMA model.

b) Models with State Space

They are described by equations of the type

$$s(t+1) = Fs(t) + Gu(t) + w(t), \quad X(t) = Hs(t) + e(t)$$

where $s(t)$ is a n dimensional vector called *state*, and F, G, H are matrices of adequate dimensions, $u(t)$ is the *input* (which is considered relevant in the problems of control; here we may assume to be 0), $w(t)$ is the so-called *noise of process*, made up of random variables pairwise uncorrelated with covariance matrix $E[w(t)\tilde{w}(s)] = \delta_{ts} R$, $e(t)$ is the so-called *noise of*

mausure, also made up of random variables pairwise uncorrelated, with covariance matrix $E[e(t)\tilde{e}(s)] = \delta_{ts} R_2$; between the two noises there is in general some correlation given by $E[w(t)\tilde{e}(s)] = \delta_{ts} R_{12}$. Beyond this, one assumes that $s(0)$, the *initial state*, is a random vector independent of the future terms of noise, with mean s and covariance matrix Π_0 . Observe that the model described above is general enough to include the multivariate case in which $X(t)$ has a dimension greater than 1. Even if it is possible to associate such a model to an ARMA model (eventually vectorial) and vice versa, the use of the state space is revealed more versatile and powerful.

5. The Kalman Filter Algorithm

Let return to the recursive algorithms of the preceding section 3. Box and Jenkins approach can be regarded as a special case of the most general and most powerful algorithm of the Kalman filter (cfr. Caines 1972). Kalman's algorithm, based on the description of the linear model by means of the state space, lends itself to be extended to multivariate processes, with little additional strength, differently from Box and Jenkins approach. Consequently in the following we only illustrate Kalman's algorithm. We are considering for the model (6) the problem of estimating the state vector, given $X(t)$ (and possibly $u(t)$). Let

$$\hat{s}(t) := E[s(t) | X(0), u(0), \dots, X(t-1), u(t-1)].$$

In the case the initial state and the involved noises are Gaussian, the preceding estimation is obtained by the following recursive procedure (*Kalman filter*):

$$\begin{aligned} \hat{s}(0) &:= s(0) \\ \hat{s}(t+1) &= F\hat{s}(t) + Gu(t) + K(t)[X(t) - H\hat{s}(t)]. \end{aligned}$$

The matrix $K(t)$ is called *Kalman gain* and is given by

$$K(t) := [FP(t)\tilde{H} + \tilde{R}_{12}] [HP(t)\tilde{H} + R_2]^{-1}$$

where the matrix $P(t)$ is a solution of the *Riccati equation*

$$P(0) := \Pi_0$$

$$P(t+1) = FP(t) \tilde{F} + R_1 - [FP(t) \tilde{H} + \tilde{R}_{12}] [HP(t) \tilde{H} + R_2]^{-1} [FP(t) \tilde{H} + \tilde{R}_{12}].$$

In general the algorithm, when the disturbances are not Gaussian, does not furnish an estimation coinciding with the conditional expectation (7); but a minimum covariance estimation among those which are linear in X and u . Finally we can note that Kalman filter works also in the most general cases where all matrices are time-dependent; but a detailed study of this would be beyond the imposed limits of the present paper.

Bibliography

- Box G.E.P. e G.M. Jenkins (1970), *Time series Analysis, Forecasting and Control*, Holden-Day, San Francisco.
- Caines P.E. (1972), *Relationship between Box-Kenkins-Astrom control and Kalman linear regulator*, Proc. IEEE, **119**, 615--620.
- Kalman R.E. (1960), *A new approach to linear filtering and prediction problems*, Trans. ASME J. Basic Engrg., ser. D, **82**, 35--45.
- Kalman R.E. e R.S. Bucy (1961), *New results in linear filtering and prediction problems*, Trans. ASME J. Basic Engrg., ser. D, **83**, 95--108.
- Kolmogoroff A.N. (1941), *Interpolation und extrapolation von stationaren Zufalligen Folgen*, Bull. Acad. Sci. (Nauk), U.S.S.R., ser. math., **5**, 3--14.
- Papoulis A. (1965), *Probability, random variables and stochastic processes*, Mc Graw Hill, New York.
- Wiener N. (1949), *The extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications*, Wiley, New York.