

## 6. Wiener and Kalman Filters

**6.1. The Wiener Filter.** The theory of filtering of stationary time series for a variety of purposes was constructed by Norbert Wiener in the 1940s for continuous time processes in a notable feat of mathematics (Wiener, 1949). The work was done much earlier, but was classified until well after World War II). In an important paper however, Levinson (1947) showed that in discrete time, the entire theory could be reduced to least squares and so was mathematically very simple. This approach is the one used here. Note that the vector space method sketched above is fully equivalent too.

The theory of Wiener filters is directed at operators (filters) which are causal. That is, they operate only upon the past and present of the time series. This requirement is essential if one is interested in forecasting so that the future is unavailable. (When the future is available, one has a “smoothing” problem.) The immediate generator of Wiener’s theory was the need during the Second World War for determining where to aim anti-aircraft guns at dodging airplanes. A simple (analogue) computer in the gunsight could track the moving airplane, thus generating a time history of its movements and some rough autocovariances. Where should the gun be aimed so that the shell arrives at the position of the airplane with smallest error? Clearly this is a forecasting problem. In the continuous time formulation, the requirement of causality leads to the need to solve a so-called Wiener-Hopf problem, and which can be mathematically tricky. No such issue arises in the discrete version, unless one seeks to solve the problem in the frequency domain where it reduces to the spectral factorization problem alluded to in Chapter 1.

### The Wiener Predictor

Consider a time series  $x_m$  with autocovariance  $R_{xx}(\tau)$  (either from theory or from prior observations). We assume that  $x_m, x_{m-1}, \dots$ , are available; that is, that enough of the past is available for practical purposes (formally the infinite past was assumed in the theory, but in practice, we do not, and cannot use, filters of infinite length). We wish to forecast one time step into the future, that is, seek a filter to construct

$$\tilde{x}_{m+1} = a_0 x_m + a_1 x_{m-1} + \dots + a_M x_{m-M} = \sum_{k=0}^M a_k x_{m-k}, \quad (6.1)$$

such that the  $a_i$  are fixed. Notice that the causality of  $a_k$  permits it only to work on present (time  $m$ ) and past values of  $x_p$ . We now minimize the “prediction error” for all times:

$$J = \sum_m (\tilde{x}_{m+1} - x_{m+1})^2. \quad (6.2)$$

This is the same problem as the one leading to (1.14) with the same solution. The prediction error is just

$$\begin{aligned} P &= \langle (\tilde{x}_{m+1} - x_{m+1})^2 \rangle \\ &= R(0) - 2 \sum_{k=0}^M a_k R(k+1) + \sum_{k=1}^M \sum_{l=1}^M a_k a_l R(k-l) \leq R(0). \end{aligned} \quad (6.3)$$

Notice that if  $x_m$  is a white noise process,  $R(\tau) = \sigma_{xx}^2 \delta_{\tau 0}$ ,  $a_i = 0$ , and the prediction error is  $\sigma_{xx}^2$ . That is to say, the best prediction one can make is  $\tilde{x}_{m+1} = 0$  and Eq. (6.3) reduces to  $P = R(0)$ , the full variance of the process and there is no prediction skill at all. These ideas were applied by Wunsch (1999) to the question of whether one could predict the NAO index with any skill through linear means. The estimated autocovariance of the NAO is shown in Fig. 1 (as is the corresponding spectral density). The conclusion was that the autocovariance is so close to that of white noise (the spectrum is nearly flat), that while there was a very slight degree of prediction skill possible, it was unlikely to be of any real interest. (The NAO is almost white noise.) Colored processes can be predicted with a skill depending directly upon how much structure their spectra have.

Serious attempts to forecast the weather by Wiener methods were made during the 1950s. They ultimately foundered with the recognition that the atmosphere has an almost white noise spectrum for periods exceeding a few days. The conclusion is not rigidly true, but is close enough that what linear predictive skill could be available is too small to be of practical use, and most sensible meteorologists abandoned these methods in favor of numerical weather prediction (which however, is still limited for related reasons, to skillful forecasts of no more than a few days). It is possible that spatial structures within the atmosphere, possibly obtainable by wavenumber filtering, would have a greater linear prediction possibility. This may well be true, but they would therefore contain only a fraction of the weather variance, and one again confronts the issue of significance. To the extent that the system is highly non-linear (non-Gaussian), it is possible that a non-linear filter could do better than a Wiener one. It is possible to show however, that for a Gaussian process, no non-linear filter can do any better than the Wiener one.

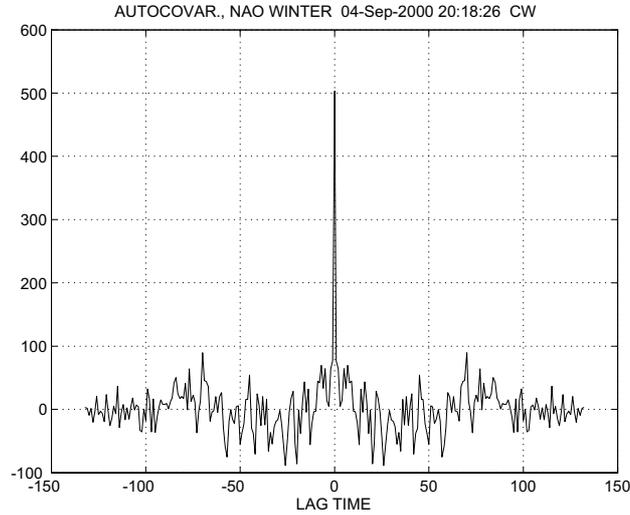


FIGURE 1. The estimated autocovariance of the North Atlantic Oscillation Index (NAO). Visually, and quantitatively, the autocovariance is dominated by the spike at the origin, and differs little from the autocovariance of a white noise process. By the Wiener filter theory, it is nearly unpredictable. See Wunsch (1999).

*Exercise.* Find the prediction error for a forecast at  $k$ -time steps into the future.

*Exercise.* Consider a vector time series,  $\mathbf{x}_m = [h_m, g_m]^T$ , where  $\langle h_m, g_m \rangle \neq 0$ . Generalize the Wiener prediction filter to this case. Can you find a predictive decomposition?

A slightly more general version of the Wiener filter (there are a number of possibilities) is directed at the extraction of a signal from noise. Let there be a time series

$$x_m = S_m + n_m \quad (6.4)$$

where  $S_m$  is the signal which is desired, and  $n_m$  is a noise field. We suppose that  $\langle S_m, n_m \rangle = 0$  and that the respective covariances  $R_{SS}(\tau) = \langle S_m, S_{m+\tau} \rangle$ ,  $R_{nn}(\tau) = \langle n_m, n_{m+\tau} \rangle$  are known, at least approximately. We seek a filter,  $a_m$ , acting on  $x_m$  so that

$$\sum_m a_m x_{k-m} \approx S_k \quad (6.5)$$

as best possible. (The range of summation has been left indefinite, as one might not always demand causality.) More formally, minimize

$$J = \sum_{k=0}^{N-1} \left( S_k - \sum_m a_m x_{k-m} \right)^2. \quad (6.6)$$

*Exercise.* Find the normal equations resulting from (6.6). If one takes  $S_m = x_{m+1}$ , is the result the same as for the prediction filter? Suppose  $a_m$  is symmetric (that is acausal), take the Fourier transform of

the normal equations, and using the Wiener-Khinchin theorem, describe how the signal extraction filter works in the frequency domain. What can you say if  $a_k$  is forced to be causal?

**6.2. The Kalman Filter.** R. Kalman (1960) in another famous paper, set out to extend Wiener filters to non-stationary processes. Here again, the immediate need was of a military nature, to forecast the trajectories of ballistic missiles, which in their launch and re-entry phases would have a very different character than a stationary process could describe. The formalism is not very much more complicated than for Wiener theory, but is best left to the references (see e.g., Wunsch, 1966, Chapter 6). But a sketch of a simplified case may perhaps give some feeling for it.

Suppose we have a “model” for calculating how  $x_m$  will behave over time. Let us assume, for simplicity, that it is just an AR(1) process

$$x_m = ax_{m-1} + \theta_m. \quad (6.7)$$

Suppose we have an estimate of  $x_{m-1}$ , called  $\tilde{x}_{m-1}$ , with an estimated error

$$P_{m-1} = \langle (\tilde{x}_{m-1} - x_{m-1})^2 \rangle. \quad (6.8)$$

Then we can make a forecast of  $x_m$ ,

$$\tilde{x}_m(-) = \tilde{a}x_{m-1} \quad (6.9)$$

because  $\theta_m$  is unknown and completely unpredictable by assumption. The minus sign in the argument indicates that no observation from time  $m$  has been used. The prediction error is now

$$P_m(-) = \langle (\tilde{x}_m(-) - x_m)^2 \rangle = \sigma_\theta^2 + P_{m-1}, \quad (6.10)$$

that is, the initial error propagates forward in time and is additive to the new error from the unknown  $\theta_m$ . Now let us further assume that we have a measurement of  $x_m$  but one which has noise in it,

$$y_m = Ex_m + \varepsilon_m, \quad (6.11)$$

where  $\langle \varepsilon_m \rangle = 0$ ,  $\langle \varepsilon_m^2 \rangle = \sigma_\varepsilon^2$ , which produces an estimate  $y_m/E$ , with error variance  $E^{-2}\sigma_\varepsilon^2$ . The observation of  $x_m$  ought to permit us to improve upon our forecast of it,  $\tilde{x}_m(-)$ . A plausible idea is to *average the measurement with the forecast*, weighting the two inversely as their relative errors:

$$\tilde{x}_m = \frac{\sigma_\theta^2 + P_{m-1}}{(\sigma_\theta^2 + P_{m-1}) + E^{-2}\sigma_\varepsilon^2} E^{-1}y_m + \frac{E^{-2}\sigma_\varepsilon^2}{(\sigma_\theta^2 + P_{m-1}) + E^{-2}\sigma_\varepsilon^2} \tilde{x}_m(-) \quad (6.12)$$

$$= \tilde{x}_m(-) + \frac{(\sigma_\theta^2 + P_{m-1}) E^{-1}}{(\sigma_\theta^2 + P_{m-1}) + E^{-2}\sigma_\varepsilon^2} (y_m - E\tilde{x}_m(-)) \quad (6.13)$$

(See Wunsch, 1996, section 3.7). If the new data are very poor relative to the forecast,  $\sigma_\varepsilon^2 \rightarrow \infty$ , the estimate reduces to the forecast. In the opposite limit when  $(\sigma_\theta^2 + P_{m-1}) \gg E^{-2}\sigma_\varepsilon^2$ , the new data give a much better estimate, and it can be confirmed that  $\tilde{x}_m \rightarrow y_m/E$  as is also sensible.

The expected error of the average is

$$P_m = \left[ (E^{-2}\sigma_\varepsilon^2)^{-1} + (\sigma_\theta^2 + P_{m-1})^{-1} \right]^{-1} \quad (6.14)$$

$$= (\sigma_\theta^2 + P_m) - (\sigma_\theta^2 + P_{m-1})^2 \left[ (\sigma_\theta^2 + P_{m-1}) + E^{-2}\sigma_\varepsilon^2 \right]^{-1}, \quad (6.15)$$

which should be studied in the two above limits. Now we proceed another time-step into the future. The new best forecast is,

$$\tilde{x}_{m+1} = a\tilde{x}_m \quad (6.16)$$

where  $\tilde{x}_m$  has error  $P_m$  and we proceed just as before, with  $m \rightarrow m + 1$ . If there are no observations available at time  $m$ , then the forecast cannot be improved,  $\tilde{x}_{m+1} = a\tilde{x}_m$ , and one keeps going with  $\tilde{x}_{m+2} = a\tilde{x}_{m+1}$ . The Kalman filter permits one to employ whatever information is contained in a model (perhaps dynamical, as in orbital equations or an ocean circulation model) along with any observations of the elements of the system that come in through time. The idea is extremely powerful and many thousands of papers and books have been written on it and its generalizations.<sup>1</sup> If there is a steady data stream, and the model satisfies certain requirements, one can show that the Kalman filter asymptotically reduces to the Wiener filter—an important conclusion because the Kalman formalism is generally computationally much more burdensome than is the Wiener one. The above derivation contains the essence of the filter, the only changes required for the more general case being the replacement of the scalar state  $x(t)$  by a vector state,  $\mathbf{x}(t)$ , and with the covariances becoming matrices whose reciprocals are inverses and one must keep track of the order of operations.

**6.3. Wiener Smoother.** “Filtering” in the technical sense involves using the present and past, perhaps infinitely far back, of a time-series so as to produce a best estimate of the value of a signal or to predict the time series. In this situation, the future values of  $x_m$  are unavailable. In most oceanographic problems however, we have a stored time series and the formal future is available. It is unsurprising that one can often make better estimates using future values than if they are unavailable (just as interpolation is more accurate than extrapolation). When the filtering problem is recast so as to employ both past and future values, one is doing “smoothing”. Formally, in equations such as (6.5, 6.6), one permits the index  $m$  to take on negative values and finds the new normal equations.

---

<sup>1</sup>Students may be interested to know that it widely rumored that Kalman twice failed his MIT general exams (in EECS).