

Kayo Ide \*and Didier Sornette  
UCLA, Los Angeles, California

## 1 INTRODUCTION

The Kalman filter provides the optimal resolution to the problem of sequential data-assimilation under the hypothesis that the true system variables evolve according to linear dynamics subject to additive normal noises, and that the observed variables are linearly related to the true system variable subject to additive normal noises.

Two main limitations that restrict the performance of the Kalman filter are therefore: 1) the nonlinearity of the real system dynamics; 2) the non-normality (non-Gaussianity) of the noises. The first item has been addressed partly by the so-called “extended” Kalman filters that performs local linearizations. This study focuses on the second item. Performance of the standard Kalman filtering is restricted especially when the noise distributions have heavy tails corresponding to extreme bursts of the system variables and observations. We present a new theoretical framework for the optimal sequential data assimilation, the Kalman-Lévy(KL) filter, when the noises are distributed according to the power or Lévy laws.

## 2 FUNDAMENTAL PROPERTIES OF POWER- AND LÉVY-LAW NOISES

### 2.1 One-dimensional noise distribution

For one-dimensional system, a power-law probability density function can be expressed as:

$$P(\omega)d\omega \simeq \frac{1}{|\omega/C^{\frac{1}{\mu}}|^{1+\mu}} d\left(\omega/C^{\frac{1}{\mu}}\right), \quad (1)$$

as  $|\omega| \rightarrow \pm\infty$ . A family of power and Lévy laws is characterized by two parameters, the exponent  $\mu$  and the scale factor  $C$ .

The exponent  $\mu$  controls the decay rate of the probability, i.e., the smaller it is, the wilder can be the noise. For  $\mu \leq 2$  (resp.  $\mu \leq 1$ ), the variance (resp. mean) is not defined mathematically. Therefore, significant light may be shed onto the data-assimilation

systems using the power- or Lévy-law for the noise distributions, instead of the Gaussian law. The scale factor  $C$  controls the global amplitude of the power law tail, i.e., the larger it is, the more frequent the extreme bursts can be. Together,  $C^{\frac{1}{\mu}}$  defines the characteristic scale of the self-similar fluctuations of  $\omega$ . For a given  $\mu$ , the uncertainty of  $\omega$  is uniquely determined by  $C$ .

We call such  $\omega$  the “ $\mu$ -variable.” It has useful properties (Sornette 2001), including:

- (a) if  $\omega$  is a  $\mu$ -variable with the scale factor  $C$ , then so is  $p\omega$  with the scale factor  $|p|^{\mu}C$  for a real  $p$ ;
- (b) if  $\omega_i$  and  $\omega_j$  are two independent  $\mu$ -variables with the scale factors  $C_i$  and  $C_j$ , then  $\omega = \omega_i + \omega_j$  is also a  $\mu$ -variable with the scale factor  $C = C_i + C_j$ .

### 2.2 Optimal estimation

We consider the estimation of an true state variable  $x^t$  using two samples  $x^f$  and  $x^o$ ; the superscripts  $\{\cdot\}^{t,f,o}$  correspond to true, forecast and observation of the data assimilation system, respectively. The two samples are contaminated by independent noises  $\omega^{f,o}$  that are both  $\mu$ -variables with the scale factor  $C^{f,o}$ , respectively. The estimate  $\hat{x}$  of  $x^t$  is sought as a linear, unbiased combination of  $x^o$  and  $x^f$  with the corresponding weights  $(1 - K^o)$  and  $K^o$ :

$$\hat{x} = (1 - K^o)x^f + K^o x^o. \quad (2)$$

Using the properties of the  $\mu$ -variables mentioned above, the scale factor of the residual error in  $\hat{x}$  is:

$$\hat{C} = (1 - K^o)^{\mu} C^f + (K^o)^{\mu} C^o. \quad (3)$$

The optimal choice of the weight  $K^o$  is defined so as to minimize the uncertainty of  $\hat{x}$ , i.e., to minimize  $C$ , i.e.,  $\frac{\partial}{\partial K^o} \hat{C} = 0$  subject to  $\frac{\partial^2}{\partial (K^o)^2} \hat{C} > 0$ . For  $\mu > 1$ , the solution for the optimal  $K^o$  exists:

$$K^o = \frac{1}{1 + \lambda^{\frac{\mu}{\mu-1}}}, \quad (4)$$

where

$$\lambda \equiv \frac{(C^o)^{\frac{1}{\mu}}}{(C^f)^{\frac{1}{\mu}}} \quad (5)$$

\*Corresponding author address: Kayo Ide, UCLA, Dept. of Atmospheric Sciences, Los Angeles, CA 90095-1565; e-mail: [kayo@atmos.ucla.edu](mailto:kayo@atmos.ucla.edu)

is the ratio of the characteristic error size of  $x^{f,o}$ . The resulting optimal scale factor is

$$\hat{C} = \frac{\lambda^\mu}{\left(1 + \lambda^{\frac{\mu}{\mu-1}}\right)^{\mu-1}} C^f. \quad (6)$$

For  $\mu = 2$ , this procedure is identical to the optimal estimation for a system where  $\omega^{f,o}$  are distributed according to the Gaussian law with the variances  $C^{f,o}$ . For  $\mu \leq 1$ , the noises are so wild that there is no gain in combining the two samples  $x^{f,o}$ . The one with the smaller scale factor must be chosen for such a case.

### 2.3 $N$ -dimensional $\mu$ -variable

For  $N$ -dimensional  $\mu$ -vector  $\epsilon$ , we rewrite it as a linear sum of  $N$  independent  $\mu$ -variables:

$$\epsilon = \mathbf{G}\omega \quad (7)$$

where  $\omega$  is the vector representation of  $N$  independent  $\mu$ -variables and  $\mathbf{G} \in \mathbb{R}^{N \times N}$  represents the linear relation between  $\epsilon$  and  $\omega$ . Then,  $\epsilon$  is characterized by two parameters, the exponent  $\mu$  and the ‘‘tail covariance matrix’’:

$$\mathbf{B} = \mathbf{G}^{[\frac{\mu}{2}]} \mathbf{C} \mathbf{G}^{T[\frac{\mu}{2}]} . \quad (8)$$

where  $\mathbf{C} \in \mathbb{R}^{N \times N}$  is a diagonal matrix associated with the individual scale factors of  $\omega$ . The operator  $\{\cdot\}^{[\beta]}$  means that each element of matrix or vector is defined by

$$\mathbf{G}_{ij}^{[\beta]} = \text{sign}(G_{ij}) |G_{ij}|^\beta . \quad (9)$$

This definition of  $\mathbf{B}$  is consistent with the properties of  $\mu$ -variables mentioned above. Like the scale factor  $C$  for  $\omega$  in the one-dimensional case, the tail covariance matrix  $\mathbf{B}$  represents the uncertainty of  $\omega$ . We measure the global uncertainty of  $N$ -dimensional  $\mu$ -variable by the ‘‘average’’ scale factor, i.e.,  $\text{trace}\mathbf{B}$ . The larger  $\text{trace}\mathbf{B}$  is, the more uncertain is  $\omega$ . For  $\mu = 2$ , the tail-covariance matrix  $\mathbf{B}$  is identical to the error covariance matrix.

## 3 THE KALMAN-LÉVY FILTER

### 3.1 Problem

We are now in a position to formulate a sequential data-assimilation framework for a system with the noises distributed according to the power or Lévy law. We present here the framework for a state vector  $\mathbf{x} \in \mathbb{R}^N$  whose true dynamics is given by a set of linear stochastic difference equations:

$$\mathbf{x}_k^t = \mathbf{M}_{k,k-1} \mathbf{x}_{k-1}^t + \boldsymbol{\eta}_{k-1}^t, \quad (10)$$

where  $\mathbf{M}_{k,k-1} \in \mathbb{R}^{N \times N}$  is the linear dynamical map and  $\boldsymbol{\eta}_{k-1}^t$  is the  $N$ -dimensional  $\mu$ -vector dynamical noise with the tail covariance  $\mathbf{B}_{k-1}^\eta$ . It is straightforward to extend the framework for a system whose dynamics is given by a set of linear stochastic differential equations.

The index  $k$  corresponds to the time sequence when the observations  $\mathbf{y}_k^o \in \mathbb{R}^{L_k}$  are taken. The observations  $\mathbf{y}_k^o$  are assumed to be linear functions of the true state vector  $\mathbf{x}_k^t$  subject to an additive noise:

$$\mathbf{y}_k^o = \mathbf{H}_k \mathbf{x}_k^t + \epsilon_k^o, \quad (11)$$

where  $\mathbf{H}_k \in \mathbb{R}^{L_k \times N}$  is the linear observation function which can vary at each time step  $k$  and  $\epsilon_k^o$  is the  $\mu$ -vector observational noise with the tail covariance  $\mathbf{B}_k^\epsilon$ .

### 3.2 Optimal sequential data-assimilation system

**STEP 1. DYNAMIC FORECAST:** Given a set of initial conditions described by the subscript  $\{\cdot\}_{k-1}$  which are known, the forecast is performed deterministically to advance from  $k-1$  to  $k$  based on (10):

$$\mathbf{x}_k^f = \mathbf{M}_{k,k-1} \mathbf{x}_{k-1}^a. \quad (12)$$

The corresponding tail-covariance of the forecasts at time  $k$  is:

$$\mathbf{B}_k^f = (\mathbf{M}_{k,k-1} \mathbf{G}_{k-1}^a)^{[\frac{\mu}{2}]} \mathbf{C}_{k-1}^a (\mathbf{M}_{k,k-1} \mathbf{G}_{k-1}^a)^{T[\frac{\mu}{2}]} + \mathbf{B}_{k-1}^\eta, \quad (13)$$

where the definition (8,9) is used for  $\mathbf{B}_{k-1}^a$ . **STEP 2. PROBABILISTIC ANALYSIS:** Given the forecast  $\mathbf{x}_k^f$  with  $\mathbf{B}_k^f$  from Step 1 along with the observations  $\mathbf{y}_k^o$  with tail-covariance  $\mathbf{B}_k^\epsilon$ , the analysis  $\mathbf{x}_k^a$  is:

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k^L (\mathbf{y}_k^o - \mathbf{H}_k \mathbf{x}_k^f) \quad (14)$$

with the tail-covariance:

$$\mathbf{B}_k^a = (\mathbf{G}_k^f - \mathbf{K}_k^L \mathbf{H}_k \mathbf{G}_k^f)^{[\frac{\mu}{2}]} \mathbf{C}_k^f (\mathbf{G}_k^f - \mathbf{K}_k^L \mathbf{H}_k \mathbf{G}_k^f)^{T[\frac{\mu}{2}]} + (\mathbf{K}_k^L \mathbf{G}_k^\epsilon)^{[\frac{\mu}{2}]} \mathbf{C}_k^\epsilon (\mathbf{K}_k^L \mathbf{G}_k^\epsilon)^{T[\frac{\mu}{2}]}, \quad (15)$$

where the definition (8,9) is used for  $\mathbf{B}_k^f$  and  $\mathbf{B}_k^\epsilon$ .

The optimal filter  $\mathbf{K}_k^L$  is determined so as to minimize the average uncertainty of  $\mathbf{x}_k^a$ , i.e.,  $\text{trace}\mathbf{B}_k^a$ . Therefore, each element of  $\mathbf{K}_k^L$  is obtained by solving for

$$\frac{\partial}{\partial \mathbf{K}_k^L} \text{trace} \mathbf{B}_k^a = 0, \quad (16)$$

with the condition that Hessian  $\frac{\partial^2}{\partial(\mathbf{K}_k^L)^2} \text{trace} \mathbf{B}_k^a$  must be positive definite. It can be shown mathematically that (16) results in  $N$  sets of  $L_k$  nonlinear equations for  $L_k$  unknowns. Each set corresponds to minimization of  $\mathbf{B}_k^a|_{ii}$ :

$$\frac{\partial}{\partial \mathbf{K}_k^L} \mathbf{B}_k^a|_{ii} = 0, \quad (17)$$

for  $i = 1, \dots, N$  and  $j = 1, \dots, L_k$ . It can be also shown mathematically that such  $\mathbf{K}_k^L$  uniquely exists for  $\mu > 1$ . For  $\mu = 2$ , the KL filter is identical to the standard Kalman filter for the Gaussian noise distributions.

## 4 NUMERICAL EXPERIMENTS

We demonstrate the performance of the KL filter using a simple one-dimensional case where the dynamics and observations are stationary. We use  $\mu = 1.2$  for the heavy-tail distribution of the wild noises,  $(M, B^\eta) = (0.9, 1)$  for the dynamics, and  $(H, B^\epsilon) = (1, 1)$  for the observations. A realization of the true variable  $x_k^t$  is shown in Figure 1a for  $k = 1, \dots, 10000$  by a solid line. The occurrence of a few large peaks are caused by rare but extreme noise fluctuations due to the Lévy noise distribution. Based on  $x_k^t$ , a realization of  $y_k^o$  (circles) is shown in Figure 1b for  $k = 1050, \dots, 1100$ .

We first perform an identical twin experiment using the KL filter. Figure 1b shows the forecasts  $x_k^f$  (square) and analyses  $x_k^a$  (diamonds), using  $y_k^o$ . Next we perform the standard Kalman filtering using the same  $y_k^o$  assuming that both dynamical and observational noises are distributed by the Gaussian laws. For this experiment, we use  $(B^\eta)^{2/\mu}$  for dynamical error variance  $P^\eta$  and  $(B^\epsilon)^{2/\mu}$  for observational error variance  $P^\epsilon$  so that the Kalman filtering is the optimal within its definition. The result of the standard Kalman filtering is shown in Figure 1c. It appears clear by visual inspection that the optimal KL analysis  $x^a$  is much closer to  $x^t$  than the non-optimal Kalman filter's  $x^a$ , more often than not. The statistical analysis using the cumulative error distributions of the numerical experiments verifies that the performance of the KL filter in fact exceeds significantly that of the standard Kalman filter.

## 5 CONCLUSION

We have introduced the new concept of a “tail covariance” that generalizes the usual notion of the covariance for the noises distributed according to power

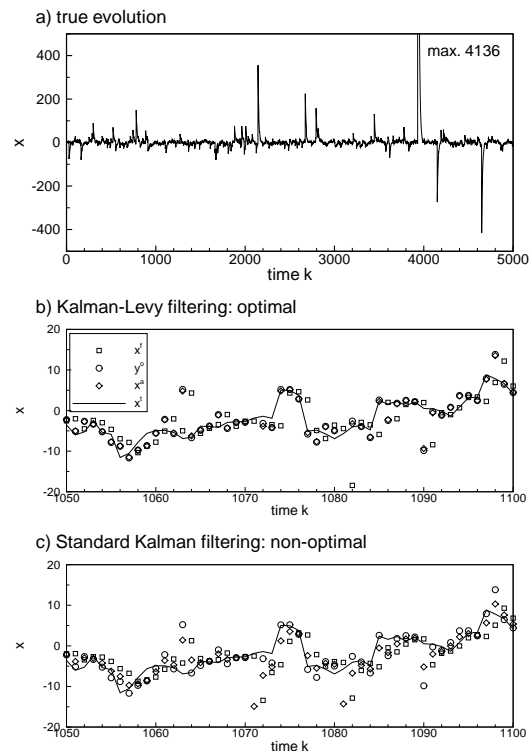


Figure 1: Numerical experiment for one-dimensional case for  $(M, B^\eta) = (0.9, 1)$  for the dynamics, and  $(H, B^\epsilon) = (1, 1)$  for observation with  $\mu = 1.2$ ; a) evolution of true state variable  $x_k^t$  for  $k = 1, \dots, 5000$ . Panel b) and c) corresponds to the use of the optimal KL filtering and the standard Kalman filter, respectively, for  $k = 1050, \dots, 1100$ .

and Lévy laws. Within the context, we have presented the exact solution of the optimal sequential data-assimilation problem. The data assimilation system is optimized so as to maximize the certainty by controlling the (heavy) tails of the error distribution. The full solution, called the Kalman-Lévy (KL) filter, is obtained as the solution of the nonlinear equations. We have investigated the performance of the KL filter in detail for a one-dimensional case and have shown by direct numerical experiments that the improvement is significant, all the more so, the heavier the tail, i.e. the smaller the power law exponent  $\mu$ .

## 6 REFERENCES

- Sornette, D., 2000: *Critical Phenomena in Natural Sciences*. Springer, 434pp.
- Sornette, D., and K. Ide, 2001: The Kalman-Lévy filter. *Physica D*, **151**, 142-174.