An Optimal Modification of a Kalman Filter for Time Scales

Charles A. Greenhall
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Dr., Pasadena, CA 91109 USA
charles.greenhall@jpl.nasa.gov

Abstract*-The Kalman filter in question, which was implemented in the time scale algorithm TA(NIST), produces time scales with poor short-term stability. A simple modification of the error covariance matrix allows the filter to produce time scales with good stability at all averaging times, as verified by simulations of clock ensembles.

I. Introduction

The purpose of a time scale is to form a virtual clock from an ensemble of physical clocks that are compared with one another at a sequence of dates, where a date is the reading of a clock as determined by counting its oscillations. The virtual clock is defined as an offset from one of the physical clocks, the offset being computed from the comparison measurements by some algorithm. The usual goal of the algorithm design is to produce a virtual clock that is more stable than any of the physical clocks in both the short term and the long term, as measured by some frequency stability measure such as Allan deviation.

A straightforward Kalman filter approach to this problem has been tried at least twice [1-3]. The noise of each clock is modeled as a sum of independent white FM and random walk FM noises with known noise levels. The entire ensemble is modeled by a linear stochastic difference equation, whose state vector is estimated by a Kalman filter from the clock comparisons, which are assumed to be noiseless. Under this assumption, if each clock's tick is offset by its Kalman phase estimate, we arrive at a single point on the time axis, called the *common corrected clocks*. It makes sense to regard this value as the estimated center of the ensemble, and to use the sequence of these values as a time scale, called the *raw Kalman scale*. This time scale, realized as TA(NIST), was reported to follow the clock with the best long-term stability, regardless of its short-term stability [4,5]. The goals of the present study have been to reproduce this result by simulation, to understand it, and to find a better way to use this Kalman filter in a time scale algorithm.

We shall exhibit two improved algorithms. The first, called the *Kalman-plus-weights scale*, ignores the Kalman phase estimates and uses the frequency state estimates in a weighted-average time scale. The second algorithm, called the *reduced Kalman scale*, uses the corrected clocks, but with a modified version of the Kalman error covariance matrix. This scale turns out to be optimal in a sense that is explained in Section VII. When applied to ensembles of dissimilar clocks, both improved scales are observed to combine the best short-term behavior with the best long-term behavior [6-8]. These results also hold for clocks with random run FM noise (random walk of drift) in addition to the other noises [7,8].

II. Ensemble Model and Kalman Filter

Let t be one of a sequence of measurement dates. When clock i ($i = 1,\ldots,n$) reaches date t internally, it issues a pulse at time $h_i(t)$, which cannot be measured by itself; instead, we can only measure the interval between the pulses of any two clocks; thus, we have the measurements $x_{ij}(t) = h_j(t) - h_i(t)$, assumed to be noiseless. The phase residual of clock i is defined by $x_i(t) = t - h_i(t)$. The sign reversal is needed: if clock i speeds up, then $h_i(t)$ tends to decrease. (Strictly speaking, we ought to write $x_i(h_i(t))$ instead of $x_i(t)$ because $h_i(t)$ is the physical time, but the difference is negligible.)

The stochastic model for clock i is given by

$$\frac{dx_i(t)}{dt} = y_i(t) + n_{xi}(t), \quad \frac{dy_i(t)}{dt} = n_{yi}(t), \tag{1}$$

where n_{xi} and n_{yi} are white noises with spectral densities $2q_{xi}$ and $2q_{yi}$, which are assumed to be known. The frequency *state* y_i contains only the random walk FM portion of the total frequency dx_i / dt.

By solving (1) for each clock between two measurement dates $t - \tau_0$ and t, we obtain the discrete-time model given by Jones and Tryon [2]:

$$X(t) = \Phi(\tau_0) X(t - \tau_0) + W(t), \tag{2}$$

where $X = [x_1, y_1, \ldots, x_n, y_n]^{\mathrm{T}}$ is the system state, $\Phi(\tau_0)$ is a $2n \times 2n$ matrix with diagonal blocks $\begin{bmatrix} 1 & \tau_0 \\ 0 & 1 \end{bmatrix}$, and the process noise $W = [w_{x1}, w_{y1}, \ldots, w_{xn}, w_{yn}]^{\mathrm{T}}$ is a vector of mean-0 random variables whose covariance matrix $Q(\tau_0)$ has diagonal blocks

* This research was conducted at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

$$\begin{bmatrix} q_{xi}\tau_0 + q_{yi}\tau_0^3/3 & q_{yi}\tau_0^2/2 \\ q_{yi}\tau_0^2/2 & q_{yi}\tau_0 \end{bmatrix}. \tag{3}$$

A set of $n-1$ independent noiseless measurements is expressed by $HX = [x_2 - x_1, \ldots, x_n - x_1]^{\mathrm{T}}$. For three clocks, $H = \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$. The measurement dates may be unequally spaced, i.e., τ_0 may depend on t.

Fig. 1 shows an idealized model of an ensemble of four clocks: two hydrogen masers and two mercury ion traps. The ion traps are assumed to have only white FM noise. Plotted are the theoretical Allan deviations $\sigma_y(\tau) = \sqrt{q_{xi}/\tau + q_{yi}\tau/3}$ (lines) and the empirical Allan deviations (dots) from a 1000-point simulation of (2) with $\tau_0 = 1000$ s.

The Kalman filter is a set of equations that propagate an unbiased minimal-variance state estimate from one measurement date to the next [9-11]. We want to obtain the estimate $\hat{X}(t)$ and its error covariance matrix

$$P(t) = \mathrm{E}\left[X(t) - \hat{X}(t)\right]\left[X(t) - \hat{X}(t)\right]^{\mathrm{T}}$$

from $\hat{X}(t-\tau_0)$ and $P(t-\tau_0)$. First we perform the temporal update (prediction) equations:

$$\begin{aligned} \tilde{X}(t) &= \Phi(\tau_0)\hat{X}(t-\tau_0), \\ \tilde{P}(t) &= \Phi(\tau_0)P(t-\tau_0)\Phi(\tau_0)^{\mathrm{T}} + Q(\tau_0). \end{aligned} \tag{4}$$

After making the noiseless measurement $HX(t)$ we perform the measurement update equations:

$$\left.\begin{aligned} K(t) &= \tilde{P}(t)H^{\mathrm{T}}\left[H\tilde{P}(t)H^{\mathrm{T}}\right]^{-1}, \\ \hat{X}(t) &= \tilde{X}(t) + K(t)\left[HX(t) - H\tilde{X}(t)\right], \\ \hat{P}(t) &= \tilde{P}(t) - \tilde{P}(t)H^{\mathrm{T}}K(t)^{\mathrm{T}}, \end{aligned}\right\} \tag{5}$$

thus completing the iteration. The $2n \times (n-1)$ matrix $K(t)$ is called the Kalman gain.

The Kalman filter must be started with some initial $\hat{X}(t)$ and $P(t)$; we shall say more about this in Section V.

For the sake of numerical stability over many iterations, it is better to propagate a Cholesky factor of the covariance matrix; this is an upper triangular matrix $C(t)$ such that $\hat{P}(t) = C(t)C(t)^{\mathrm{T}}$. There are several "square root filter" algorithms, algebraically equivalent to (4) and (5), that propagate $\hat{X}(t)$ and $C(t)$ [10,11]. The author has used one of these methods for all simulations with no trouble.

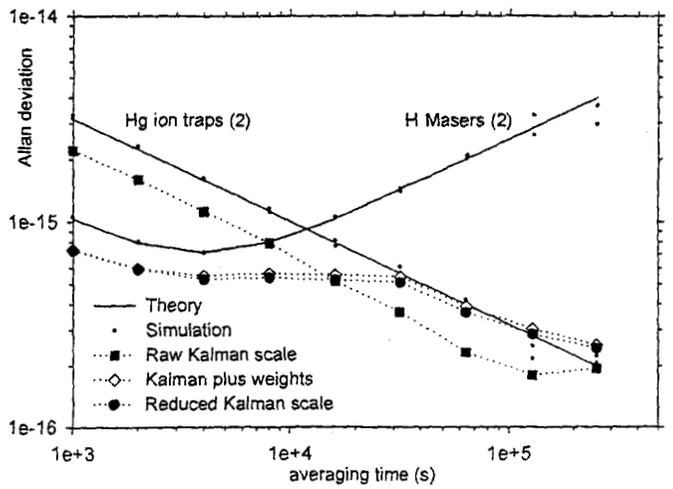

Fig. 1. Stability results from simulating an ensemble of four idealized clocks.

III. Raw Kalman Scale

From (5) we find that

$$HK(t) = I, \quad H\hat{X}(t) = HX(t), \quad HP(t) = 0. \tag{6}$$

Thus, for noiseless measurements, the estimated state satisfies the measurement equations. For the clock ensemble,

$$\hat{x}_i(t) - \hat{x}_1(t) = x_i(t) - x_1(t), \quad i = 2, \ldots, n. \tag{7}$$

It follows that the quantities

$$x_{\mathrm{e}}(t) = x_i(t) - \hat{x}_i(t) \text{ and } h_{\mathrm{e}}(t) = h_i(t) + \hat{x}_i(t) \tag{8}$$

do not depend on i. Brown [3] called $h_i(t) + \hat{x}_i(t)$ the *corrected clocks*; for convenience we shall use this term for the quantities $x_i(t) - \hat{x}_i(t)$, whose common value $x_{\mathrm{e}}(t)$ is called the *raw Kalman scale*, or Kraw.

The squares in Fig. 1 show the Allan deviation of Kraw for the simulated clocks. We find that $Kraw(t)$ ignores the masers and converges quickly to the mean of the two ion trap phases. Although the long-term stability of Kraw is as good as it can be, it does not take advantage of the quieter clocks to improve its short-term stability.

IV. Kalman Plus Weights

To improve this situation, one can ignore the phase estimates altogether and use the Kalman frequency state estimates in a weighted-average time scale governed by the *basic time scale equation* (BTSE) ([12], p. 133 ff.). The BTSE has several equivalent forms; the one used here is a sequential definition of the time scale (denoted again by $x_{\mathrm{e}}(t)$) in terms of the non-observable quantities $x_i(t)$:

$$\begin{aligned} & x_{\mathrm{e}}(t) - x_{\mathrm{e}}(t-\tau_0) \\ & = \sum_{i=1}^{n} \lambda_i(t)\left[x_i(t) - x_i(t-\tau_0) - \tau_0\hat{y}_i(t-\tau_0)\right]. \end{aligned} \tag{9}$$

The weights $\lambda_i(t)$ (to be determined) add to 1. One can convert (9) to a more conventional form of the BTSE,

$$x_{ej}(t)=\sum_{i=1}^{n}\lambda_i(t)\left[x_{ij}(t)+x_{ei}(t-\tau_0)-\tau_0\hat{y}_i(t-\tau_0)\right],$$

which gives the ensemble offsets $x_{ej}(t)=x_e(t)-x_j(t)$ sequentially from the measurements $x_{ij}(t)=x_i(t)-x_j(t)$.

The presence of Kalman frequency estimates in (9) is unconventional; the frequency estimates for weighted-average scales are usually obtained by estimating the frequency departure of each clock from the scale itself as it evolves. Here, we feed no information from the time scale back to the Kalman filter, whose only job at this point is to deliver frequency estimates from the clock models and measurements.

To select the weights we observe that the *i*th bracketed term of (9) is the τ_0-increment of a detrended version of the phase of clock *i*. Because all these detrended phases have good long-term stability, we may assign the weights to optimize their combined short-term stability. By (2) and (9),

$$\begin{aligned} &x_e(t)-x_e(t-\tau_0) \\ &=\sum_{i=1}^{n}\lambda_i(t)\left\{\tau_0\left[y_i(t-\tau_0)-\hat{y}_i(t-\tau_0)\right]+w_{xi}(t)\right\}. \end{aligned} \tag{10}$$

For now, assuming that $w_{xi}(t)$ dominates the other terms in the braces, we make $\lambda_i(t)$ inversely proportional to

$$\mathrm{E}\,w_{xi}^2(t)=q_{xi}\tau_0+q_{yi}\tau_0^3/3$$

from (3). The resulting time scale is called the *Kalman-plus-weights scale*, or KPW. In Section VII we shall obtain the optimal weights without any approximation.

The open diamonds in Fig. 1 show the measured Allan deviation of KPW for our four-clock simulation. In the short term it has the stability of the average of the two masers; in the long term it has the stability of one ion trap. But for a lucky discovery, the story would have ended here.

V. Covariance X-Reduction

By an *x*-row or *x*-column of the Kalman covariance matrix $P(t)$ we mean a row or column that corresponds to a phase component of the state vector. According to the third equation in (6), all the *x*-rows (and hence all the *x*-columns) of $P(t)$ are the same vector after a measurement update; let us call it the *x-vector*. This is another way of saying that we have perfect knowledge of the clocks with respect to each other, that is, the sets $\{x_i(t), i=0,\ldots,n\}$ and $\{\hat{x}_i(t), i=0,\ldots,n\}$ differ only by an unknown rigid translation. All the phase error variances (the *x*-components of the *x*-vector) are the same. When we run the Kalman filter, we observe that this common error variance diverges fast; the filter seems to know that it is doing a poor job on the phases. On the other hand, the submatrix P_y of the frequency error covariances is empirically well behaved, though a slower divergence is not ruled out.

Suppose, now, that we declare arbitrarily that the $\hat{x}_i(t)$ are the correct phases. Then their error variances become zero, and hence the troublesome x-vector becomes zero. *Covariance x-reduction* consists in setting all x-rows and x-columns of $P(t)$ to zero. For two clocks, the operation looks like this:

P	x_1	y_1	x_2	y_2
x_1	*	*	*	*
y_1	*	*	*	*
x_2	*	*	*	*
y_2	*	*	*	*

$\rightarrow$

P'	x_1	y_1	x_2	y_2
x_1	0	0	0	0
y_1	0	*	0	*
x_2	0	0	0	0
y_2	0	*	0	*

;

the only remaining nonzero elements are in the covariance submatrix $P_y(t)$ of the frequency states.

What does this crude operation do to the Kalman filter estimates? Here is one part of the answer.

THEOREM 1. *Covariance x-reduction is transparent to all subsequent Kalman frequency estimates and their error covariances.*

In other words, we may perform the x-reduction at any date without affecting any future frequency estimates. A proof [8] is carried out by mapping the original X model to one with the state vector

$$Y=\left[0,y_1,x_2-x_1,y_2,\ldots,x_n-x_1,y_n\right]^{\mathrm{T}}.$$

By following the Kalman filter equations (4) and (5), one shows that a temporal or measurement update of the X model, with and without x-reduction, maps to the same update of the Y model.

The author first used this result for initializing the Kalman filter as follows: we run the covariance iteration without any measurement data until the P_y submatrix settles down, then use the x-reduced P as the initial error covariance; it is reasonable to treat the initial clock phases as known. For the simulations, initial frequency state errors $y_i-\hat{y}_i$ were generated as Gaussian random variables with covariance P_y.

Another use of x-reduction is to control the growth of $P(t)$, if only to improve the numerical stability of the Kalman algorithm. Brown [3] proposed a more sophisticated method, called transparent variations of covariance, for reducing $P(t)$ in a way that preserves future estimates of the entire state vector. When running the KPW scale, one can use the simpler method of x-reduction because only the frequency state estimates matter.

VI. Reduced Kalman Scale

X-reduction is not transparent to the phase estimates. The author tried it in a simulation, performing the x-reduction at each date. To his surprise, the time scale defined by the common corrected clocks (8) no longer resembled the the raw Kalman scale; it became at least as stable as the KPW scale. To repeat, one carries out the Kalman filter as before, but performs covariance x-reduction after each measurement update. Then the common corrected clocks define a new time scale, called the *reduced Kalman scale*, or Kred. The circles

in Fig. 1 show the measured Allan deviation of Kred for our simulation; it is a slight improvement over KPW over the whole range of averaging times. Other simulations gave similar results. The next section tries to explain this surprising empirical finding.

VII. The Implicit Weights

Weiss, Allan, and Peppler [4] observed that the scale defined by the corrected clocks is actually a weighted-average scale, with weights that come directly from the Kalman gain matrix. Let us see how this works. The first (x_1) component of the equation for $\hat{X}(t)$ in (5) can be written as

$$\hat{x}_1(t) = \tilde{x}_1(t) + \sum_{i=2}^{n} K_{1i}(t)\left[x_i(t) - x_1(t) - \tilde{x}_i(t) + \tilde{x}_1(t)\right], \quad (11)$$

where $\left[K_{1i}(t), i = 2, \ldots n\right]$ is the first row of $K(t)$, connecting the phase of clock 1 to the $n-1$ measurements. The *implicit weights* are defined by

$$\lambda_1(t) = 1 + \sum_{i=2}^{n} K_{1i}(t); \quad \lambda_i(t) = -K_{1i}(t), \; i = 2, \ldots, n. \quad (12)$$

These weights add to 1, but it is possible for some of them to be negative. The same weights can be obtained by similar formulas from any of the x-rows of $K(t)$. From (8), (11), and (12),

$$x_e(t) = x_1(t) - \hat{x}_1(t) = \sum_{i=1}^{n} \lambda_i(t)\left[x_i(t) - \tilde{x}_i(t)\right]. \quad (13)$$

From (4) and (8),

$$\tilde{x}_i(t) = x_i(t - \tau_0) - x_e(t - \tau_0) + \tau_0 \hat{y}_i(t - \tau_0).$$

When this is substituted into (13), the BTSE (9) appears. Let us state this finding as a theorem.

THEOREM 2. *No matter how the state estimate* $\left(\hat{X}, P\right)$ *at the previous measurement date was obtained, the common corrected clocks at the current date are obtained from the corrected clocks at the previous date by the basic time scale equation, whose weights are obtained from the Kalman gain by* (12).

This is true for both the raw and the reduced Kalman scales, but the implicit weights are different because the Kalman gains are different. Here are the implicit weights for the raw and reduced scales at the end of the four-clock simulation, together with the explicit weights of the KPW scale.

	λ_H	λ_{Hg}
Kraw	0	0.50
KPW	0.45	0.05
Kred	0.40	0.10

The BTSE from Kraw ignores the H masers entirely. The Kred weights are a little more balanced than the KPW weights, while the Kred stability is hardly better than the KPW stability.

These observations suggest that KPW and Kred are close to an optimal situation of some kind. By Theorems 1 and 2, both scales are weighted-average scales based on the same Kalman frequency estimates. It turns out that Kred itself is optimal in the following sense.

THEOREM 3. *Among all weighted-average time scales based on the Kalman frequency estimates, the reduced Kalman scale has the implicit weights that minimize the variance of the scale increment* $x_e(t) - x_e(t - \tau_0)$ *between successive measurement dates.*

This result is reasonable: by x-reducing the previous covariance matrix, we declare that the previous phase estimates are perfect; thus, when the Kalman filter optimizes the current phase estimates, it also optimizes the estimates of the phase increments.

The author has given an algebraic proof [8], which we sketch here. In fact, the optimal weights can be obtained by minimizing the variance of the expression (10) for the scale increment. The optimal solution for the vector $\left[\lambda_1, \ldots, \lambda_n\right]$ depends on $\mathrm{E}\, w_{xi}^2(\tau_0)$ and the frequency error covariance submatrix $P_y(t - \tau_0)$. On the other hand we find from (5) that $K(t)$ depends on $\mathrm{E}\, w_{xi}^2(\tau_0)$ and the entire matrix $P(t - \tau_0)$, which reduces to $P_y(t - \tau_0)$ (and zeros) if it is x-reduced. One shows that the vector of implicit weights obtained from $K(t)$ satisfies the optimality equation that comes from (10). Thus, there is no need to solve this equation for the weights; they come free with the reduced Kalman scale algorithm.

VIII. Final Remarks

As we saw, the raw Kalman scale tends to adhere to an average of the clocks that are best in the long term, regardless of their short-term behavior. According to Weiss and Weissert [5], this happens because the Kalman filter is an *accuracy* algorithm; among other things it is trying to minimize the accumulated phase estimation error over time. Their insight appears to be correct. To change this behavior, we reach inside the Kalman filter at each step to modify the error covariance matrix by the operation of x-reduction. This simple, crude operation converts the accuracy algorithm to a stability algorithm, the reduced Kalman scale. The KPW scale may be regarded as a suboptimal version of the reduced Kalman scale, with explicit weights instead of the automatic implicit weights.

This study has been carried out in a simulation playpen in which all clocks behave according to their assumed models. There is no provision for outliers, jumps in phase and frequency, changes in the ensemble, and estimation of the process noise parameters. Perhaps the simplicity and excellent stability of the reduced Kalman scale, under ideal conditions to be sure, will motivate an effort to supply these functions.

REFERENCES

[1] J. Barnes, R. Jones, P. Tryon, and D. Allan, "Stochastic models for atomic clocks," in *Proc. 14th Ann. Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, pp. 295-306, 1982.

[2] R. Jones and P. Tryon, "Continuous time series models for unequally spaced data applied to modeling atomic clocks," *SIAM J. Sci. Stat. Comput.*, vol. 8, pp. 71-81, 1987.

[3] K. Brown Jr., "The theory of the GPS composite clock," in *Proc. ION GPS-91 Meeting*, pp. 223-241, 1991.

[4] M. Weiss, D. Allan, and T. Peppler, "A study of the NBS time scale algorithm," *IEEE Trans. Instrum. Meas.*, vol. 38, pp. 631-635, 1989.

[5] M. Weiss and T. Weissert, "AT2, a new time scale algorithm: AT1 plus frequency variance," *Metrologia*, vol. 28, pp. 65-74, 1991.

[6] C. Greenhall, "Kalman plus weights: a time scale algorithm," in *Proc. 33rd Ann. Precise Time and Time Interval (PTTI) Systems and Applications Meeting*, pp. 445-454, 2001.

[7] C. Greenhall, "Addition of random run FM noise to the KPW time scale algorithm," *IV Time Scale Algorithms Symposium*, Sèvres, France, 2002, proceedings in press.

[8] C. Greenhall, "Forming stable time scales from the Jones-Tryon Kalman filter," *Metrologia*, in press.

[9] A. Gelb, ed., *Applied Optimal Estimation*, Cambridge, MA: MIT Press, 1974.

[10] G. Bierman, *Factorization Methods for Discrete Sequential Estimation*, New York: Academic Press, 1977.

[11] M. Grewal and A. Andrews, *Kalman Filtering: Theory and Practice Using Matlab*, New York: Wiley, 2001.

[12] *Selection and Use of Precise Frequency and Time Systems*, Geneva: International Telecommunication Union, 1997.