Metrologia

Forming stable timescales from the Jones–Tryon Kalman filter

To cite this article: Charles A Greenhall 2003 *Metrologia* **40** S335

INSTITUTE OF PHYSICS PUBLISHING METROLOGIA
Metrologia **40** (2003) S335–S341 PII: S0026-1394(03)61410-6

Forming stable timescales from the Jones–Tryon Kalman filter

Charles A Greenhall

Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., MS 298-100, Pasadena, CA 91109, USA

E-mail: charles.greenhall@jpl.nasa.gov

Received 24 April 2002
Published 5 June 2003
Online at stacks.iop.org/Met/40/S335

Abstract
This is a study of three timescales formed from a Kalman filter operating on a model of a clock ensemble. The raw Kalman scale is unstable at short averaging times. The Kalman-plus-weights and reduced Kalman scales are stable at all averaging times. An optimality property is proved for the reduced Kalman scale.

1. Introduction

The purpose of a timescale is to form a virtual clock from an ensemble of physical clocks whose differences from each other are measured at a sequence of dates, where a date is the displayed time 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 measurement data 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 expressed by some frequency stability measure such as Allan deviation or Hadamard deviation.

A straight Kalman filter approach to this problem has been tried at least twice [1–3]. The noise of each clock is modelled as a sum of white FM, random walk FM, and random run FM (random walk of drift), with known noise levels. The entire ensemble is modelled by a linear stochastic differential equation, whose state vector is estimated in a straightforward way by a Kalman filter from clock difference measurements. Under the assumption of noiseless measurements, if each clock's tick is offset by its Kalman phase estimate, we arrive at a single point on the time axis. It makes sense to regard this point as the estimated centre of the ensemble, and to use the sequence of these values as a timescale. This timescale, 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 this paper are to reproduce this result by simulations, to understand it, and to find a better way to use this Kalman filter in a timescale algorithm.

The model and Kalman filter are described here in detail. By simulation of an imaginary ensemble, we reproduce the reported behaviour of TA(NIST), called here the raw Kalman scale. Next, we show that a much better timescale, 'Kalman plus weights' (KPW), can be formed by using only the Kalman frequency and drift estimates in a conventional timescale equation. Finally, we describe a simple modification of the Kalman filter that turns the raw Kalman scale into the reduced Kalman scale, which has an optimal short-term stability property.

2. Jones–Tryon model and Kalman filter

The ensemble of n independent clocks has state vector

$$X = [x_1, y_1, z_1, \ldots, x_n, y_n, z_n], \tag{1}$$

where x_i is the phase, y_i the frequency state, and z_i the drift state of the ith clock. The frequency and drift states should not be confused with the total frequency $\mathrm{d}x_i/\mathrm{d}t$ and total drift $\mathrm{d}^2x_i/\mathrm{d}t^2$, which contain all three noise components. With Brown [3] we regard these states as residuals from some ideal clock whose rate is constant. The evolution of the ith clock from date $t-\delta$ to date t is described by the equations

$$\Delta_\delta x_i(t) = \delta y_i(t-\delta) + \tfrac{1}{2}\delta^2 z_i(t-\delta) + w_{xi}(t), \tag{2}$$

$$\Delta_\delta y_i(t) = \delta z_i(t-\delta) + w_{yi}(t), \tag{3}$$

$$\Delta_\delta z_i(t) = w_{zi}(t), \tag{4}$$

where Δ_δ is the backward difference operator, $\Delta_\delta f(t) = f(t) - f(t-\delta)$. The process–noise vector

$$W_i(t) = [w_{xi}(t), w_{yi}(t), w_{zi}(t)]^{\mathrm{T}}$$

has covariance matrix

$$Q_i(\delta) = \begin{bmatrix} q_{xi}\delta + \dfrac{q_{yi}\delta^3}{3} + \dfrac{q_{zi}\delta^5}{20} & * & * \\ \dfrac{q_{yi}\delta^2}{2} + \dfrac{q_{zi}\delta^4}{8} & q_{yi}\delta + \dfrac{q_{zi}\delta^3}{3} & * \\ \dfrac{q_{zi}\delta^3}{6} & \dfrac{q_{zi}\delta^2}{2} & q_{zi}\delta \end{bmatrix}. \quad (5)$$

The q factors (differential variances), which specify the levels of the three noise components for each clock, are related to the Hadamard variances of the clocks [6] by

$$\mathrm{H}\sigma^2_{yi}(\tau) = \frac{q_{xi}}{\tau} + \frac{q_{yi}\tau}{6} + \frac{11 q_{zi}\tau^3}{120}. \quad (6)$$

The noiseless clock difference measurements at date t are expressed by

$$x_{i1}(t) = x_i(t) - x_1(t), \qquad i = 2, \ldots, n. \quad (7)$$

The model is set up for Kalman filtering by expressing it as an overall matrix–vector equation

$$X(t) = \Phi(\delta)X(t-\delta) + W(t).$$

The transition matrix $\Phi(\delta)$ has diagonal blocks

$$\phi(\delta) = \begin{bmatrix} 1 & \delta & \dfrac{\delta^2}{2} \\ 0 & 1 & \delta \\ 0 & 0 & 1 \end{bmatrix}.$$

The covariance matrix of the process–noise vector $W(t) = [W_1(t), \ldots, W_n(t)]^{\mathrm{T}}$ is $Q(\delta)$, a matrix of diagonal blocks $Q_i(\delta)$. The measurement equations (7) are written as

$$\xi(t) = HX(t),$$

where $\xi(t) = [x_{21}(t), \ldots, x_{n1}(t)]^{\mathrm{T}}$, and H is an $(n-1) \times 3n$ matrix. For three clocks,

$$H = \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

The model and measurements determine a Kalman filter [7–9]. For each measurement date t, the filter produces a state estimate,

$$\hat{X}(t) = [\hat{x}_1(t), \hat{y}_1(t), \hat{z}_1(t), \ldots, \hat{x}_n(t), \hat{y}_n(t), \hat{z}_n(t)]^{\mathrm{T}},$$

and an error covariance matrix,

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

as functions of $\hat{X}(t-\delta)$, $P(t-\delta)$, and $\xi(t)$, where $t-\delta$ is the previous measurement date. The measurement dates may be unequally spaced, that is, δ may depend on t. The Kalman filter has two steps: a temporal update or prediction giving an intermediate estimate $\tilde{X}(t)$ and $\tilde{P}(t)$ from $\hat{X}(t-\delta)$ and $P(t-\delta)$, and a measurement update giving the final estimate $\hat{X}(t)$ and $P(t)$ from $\tilde{X}(t)$, $\tilde{P}(t)$, and $\xi(t)$. We assume that the reader is acquainted with Kalman filtering, whose equations will be brought up as they are needed.

2.1. Raw Kalman scale

For any Kalman filter with noiseless measurements, the estimated state satisfies the measurement equations. In fact, we have

$$\xi(t) = H\hat{X}(t), \qquad HP(t) = 0. \quad (8)$$

For our model,

$$x_{i1}(t) = \hat{x}_i(t) - \hat{x}_1(t), \qquad i = 2, \ldots, n \quad (9)$$

and the x rows and x columns of the covariance matrix are all the same. It follows from (7) and (9) that the quantity

$$x_{\mathrm{e}}(t) = x_i(t) - \hat{x}_i(t) \quad (10)$$

(which is just the phase estimate error) does not depend on i. This quantity is called the raw Kalman scale, or Kraw(t). In Brown's terminology [3], the 'corrected clocks' are all the same. It is this scale that was used for TA(NIST). As we shall see in a simulated example, the Kraw scale fails to average out the short-term noises of the clocks in the ensemble; the Kalman filter seems to attribute incorrect amounts of white FM noise to the various clocks. In section 4.3 we shall attain some insight into this situation.

2.2. Initializing the Kalman filter

The Kalman filter must be given an initial state estimate and error covariance matrix. For this paper, we make the Kalman filter itself do the work of assigning an initial error covariance by means of a preliminary filter run whose state estimates are ignored. We start the filter with a zero error covariance, and run it until the error covariance submatrix P_{yz} of the frequency and drift states settles down. It is not necessary to wait until it actually converges (if it ever does). Why just the frequency and drift states? Because the phase error variances diverge strongly; the Kalman filter seems to know that it is doing a poor job of estimating the clock phases. In fact, we may now clear all the elements of P outside the P_{yz} submatrix to zero; in section 4.1 we prove that doing so leaves the future frequency and drift state estimates unchanged. One could regard the initial P_{yz} as Type B estimates of the uncertainties [10] of the initial frequency and drift states, which might have been obtained by other means.

For the simulations in this paper, the initial frequency and drift state errors were generated as zero-mean Gaussian random variables whose covariance is the initial P_{yz} that was determined from the above procedure.

2.3. Simulation example

One of the simulation examples in a previous paper [11] on two-stage clocks reproduces an imaginary eight-clock ensemble that was simulated by Stein [12]. The odd-numbered clocks all have the same qs; similarly for the even-numbered clocks. For the purpose of this paper, a random run FM component was added to the odd-numbered clocks. A run of 1.8×10^8 s

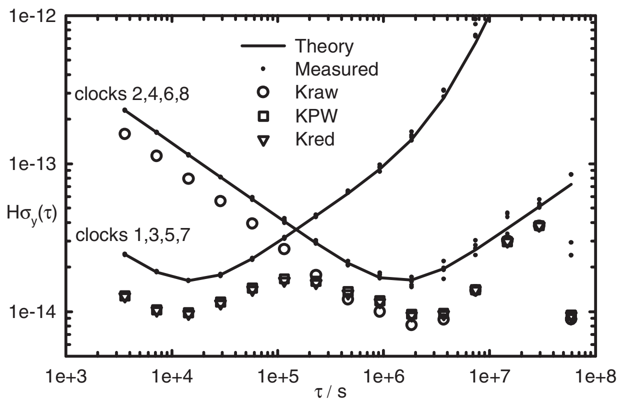

Figure 1. Results of a simulation of eight imaginary clocks. The odd-numbered clocks are statistically identical; so are the even-numbered clocks. The lines and dots show the theoretical and measured Hadamard deviations of the simulated clocks. The other symbols show the measured deviations of three timescales: the Kraw scale, the KPW scale, and the reduced Kalman scale.

of hourly measurements was simulated; the results are shown in figure 1, a Hadamard σ–τ plot. The short-term performance of the Kraw scale is scarcely better than any one of the noisiest clocks. Curiously, though, it does outperform the other scales in a range of averaging times around 10^6 s.

3. Weighted-average scales

On examining the Kalman state estimates from various simulations, one sees that the frequency and drift state estimates are of much higher quality than the phase estimates, which have incorrect amounts of short-term noise. Accordingly, the author tried an approach that uses the Kalman frequency and drift state estimates in a conventional weighted-average timescale defined by the basic timescale equation (BTSE) [13]. To include drift estimates, we use the modification introduced by Breakiron [14]. The BTSE has several equivalent forms; the one used here is a recursive definition of the timescale $x_e(t)$ in terms of the non-observable quantities $x_i(t)$:

$$\Delta_\delta x_e(t) = \sum_{i=1}^{n} \lambda_i(t) \left[\Delta_\delta x_i(t) - \delta \hat{y}_i(t-\delta) - \frac{1}{2}\delta^2 \hat{z}_i(t-\delta) \right]. \tag{11}$$

The behaviour of the scale depends on how the weights $\lambda_i(t)$ (which add to 1) are chosen, and how the estimates $\hat{y}_i(t-\delta)$ and $\hat{z}_i(t-\delta)$ are determined from previous observations. By subtracting $x_1(t)$ from both sides of (11), we obtain a recursion for the offset of the scale from clock 1 in terms of observed and computed quantities.

Here, we use the Kalman estimates $\hat{y}_i(t-\delta)$ and $\hat{z}_i(t-\delta)$ in (11). This differs from the customary practice of using estimated departures of the frequency and drift of the ith clock from the previously computed timescale. As Guinot [15] pointed out, this practice can result in ineffective frequency estimates during periods when the weights are held constant. This phenomenon does not occur here. It is important to observe that information from the BTSE is not fed back to the Kalman filter, which is kept as pure as possible; its only job at this point is to deliver frequency and drift state estimates from the clock models and measurements. The Kalman phase estimates are ignored; in section 4, however, we shall modify the Kalman filter in such a way that the phase estimates become useful.

3.1. KPW scale

It remains to determine the weights $\lambda_i(t)$ in the BTSE. We do so with the intent of minimizing the short-term instability of $x_e(t)$. By (2) and (11),

$$\Delta_\delta x_e(t) = \sum_{i=1}^{n} \lambda_i(t) \big\{ \delta[y_i(t-\delta) - \hat{y}_i(t-\delta)] + \tfrac{1}{2}\delta^2 [z_i(t-\delta) - \hat{z}_i(t-\delta)] + w_{xi}(t) \big\}. \tag{12}$$

For now, we assume that $w_{xi}(t)$ dominates the other terms in the braces. In this case, $\Delta_\delta x_e(t)$ is approximately a weighted average of the uncorrelated random variables $w_{xi}(t)$, and $x_e(t)$ is approximately a random walk in the short term. To minimize its instability, we make $\lambda_i(t)$ proportional to $1/r_i$, where

$$r_i = E w_{xi}^2(t) = q_{xi}\delta + \frac{q_{yi}\delta^3}{3} + \frac{q_{zi}\delta^5}{20} \tag{13}$$

is obtained from (5). One would expect this approximation to be best when the clocks' white FM noises dominate the other noises at the averaging time δ. In section 4.4 we shall determine the optimal weights without any approximation.

The timescale $x_e(t)$ determined by these particular weights is called the Kalman plus weights. As figure 1 shows, its measured instability in our simulation example is almost a factor of 2 below that of the best clock for all averaging times except the largest ones, where the confidence of the stability estimates is low.

4. Covariance x-reduction

The diagonal phase variance entries of the error covariance matrix $P(t)$ grow without bound, while the $(2n) \times (2n)$ submatrix $P_{yz}(t)$ of the frequency and drift covariances is empirically well behaved, though a slower divergence is not ruled out. For no other reason than to preserve the numerical stability of the Kalman algorithm, one would like to keep $P(t)$ from running away, if that can be done without disturbing the desired output of the Kalman filter. Brown [3] proposed a method, transparent variations of covariance, for reducing $P(t)$ in a way that preserves future estimates of the entire state vector. In the GPS Kalman filter, a method of pseudomeasurements has been used [16]. For constructing the KPW scale, we are interested only in preserving the frequency and drift state estimates. As we mentioned in section 2.2, there is a method, much cruder than Brown's, that accomplishes this: one simply sets all elements of $P(t)$ outside $P_{yz}(t)$ to zero. Let us call this operation *x-reduction*.

This operation is not as drastic as it may seem. As we pointed out in section 2.1, all the x rows and x columns of $P(t)$ are the same vector; we are simply setting that vector to zero.

Most of the remainder of this paper is devoted to stating and proving some remarkable properties of this simple operation.

4.1. Transparency properties

First, we must show that x-reduction is transparent to the Kalman frequency and drift state estimates. We shall obtain this result as a special case of a more general theorem.

The method uses an auxiliary model. Given the original state vector X from (1), we define the new state vector

$$Y = [\xi_1, y_1, z_1, \xi_2, y_2, z_2, \ldots, \xi_n, y_n, z_n]^{\mathrm{T}},$$

where $\xi_i = x_i - x_1$. (In particular, $\xi_1 = 0$; this convention preserves the indexing and permits a more general measurement matrix.) We can write $Y = MX$ for a certain $(3n) \times (3n)$ matrix M, which we shall not take the space to write here. The evolution equations for y_i and z_i are the same as before. The ξ_i evolve according to the equation

$$\Delta_\delta \xi_i(t) = \delta[y_i(t-\delta) - y_1(t-\delta)] + \tfrac{1}{2}\delta^2[z_i(t-\delta) - z_1(t-\delta)] + w_{xi}(t) - w_{x1}(t).$$

The measurements are just the state components $\xi_2, \ldots, \xi_n$. Thus, the new state vector $Y(t)$ can be estimated by a Kalman filter.

This model differs from the master clock relativization model discussed by Brown [3] in that the state vector contains the individual y_i and z_i, not their offsets from y_1 and z_1.

By definition, we shall say that the X estimate $(\hat{X}, P)$ maps to the Y estimate $(\hat{Y}, P_Y)$ provided that

$$\hat{Y} = M\hat{X}, \qquad P_Y = MPM^{\mathrm{T}}. \tag{14}$$

The first equation merely says that $\hat{Y}$ has the same frequency and drift components as $\hat{X}$, and its ξ components are the phase differences of $\hat{X}$.

In the following theorem, we allow measurement noise, the same for both models.

Theorem 1. *The mapping of $(\hat{X}, P)$ to $(\hat{Y}, P_Y)$ is preserved by corresponding Kalman temporal or measurement updates on the two models.*

Proof. The Y model and its measurements can be expressed in vector–matrix form,

$$Y(t) = \Phi_Y Y(t-\delta) + W_Y, \qquad \xi(t) = H_Y Y(t) + V,$$

where $W_Y = MW$, $\mathrm{cov} W_Y = Q_Y = MQM^{\mathrm{T}}$, and the $(n-1) \times (3n)$ measurement matrix H_Y has a single 1 in each row. We allow the measurement noise V to have a covariance matrix R. The reader is invited to write down M, Φ_Y, and H_Y, and to verify the identities

$$\Phi_Y M = M\Phi, \qquad H_Y M = H. \tag{15}$$

Let $(\hat{X}, P)$ map to $(\hat{Y}, P_Y)$. Corresponding Kalman temporal updates of the two models are given by

$$\tilde{X} = \Phi\hat{X}, \qquad \tilde{Y} = \Phi_Y\hat{Y},$$

$$\tilde{P} = \Phi P \Phi^{\mathrm{T}} + Q, \qquad \tilde{P}_Y = \Phi_Y P_Y \Phi_Y^{\mathrm{T}} + Q_Y.$$

Using (14) and the first identity in (15), we find easily that

$$\tilde{Y} = M\tilde{X}, \qquad \tilde{P}_Y = M\tilde{P}M^{\mathrm{T}}. \tag{16}$$

For the measurement updates, first consider the Kalman gain 'denominators'

$$D = H\tilde{P}H^{\mathrm{T}} + R, \qquad D_Y = H_Y\tilde{P}_Y H_Y^{\mathrm{T}} + R.$$

From (16) and the second identity in (15), we find that $D_Y = D$. It can be shown that D is positive definite as long as no more than one clock is noiseless, even if $R = 0$.

Let ξ be the measurement vector for both models. The measurement updates of the two models to new state estimates and covariances are given by the following equations:

$$K = \tilde{P}H^{\mathrm{T}}D^{-1}, \qquad K_Y = \tilde{P}_Y H_Y^{\mathrm{T}} D^{-1},$$

$$\hat{X} = \tilde{X} + K(\xi - H\tilde{X}), \qquad \hat{Y} = \tilde{Y} + K_Y(\xi - H_Y\tilde{Y}),$$

$$P = \tilde{P} - KDK^{\mathrm{T}}, \qquad P_Y = \tilde{P}_Y - K_Y D K_Y^{\mathrm{T}}.$$

Again from (16) and (15), we find that $K_Y = MK$, $\hat{Y} = M\hat{X}$, $P_Y = MPM^{\mathrm{T}}$. ■

By definition, two X estimates $(\hat{X}, P)$ and $(\hat{X}', P')$ that map to the same Y estimate are said to be consistent. In particular, they have the same phase differences, frequency states, and drift states. As a corollary of theorem 1, consistent X estimates remain consistent if the same sequence of Kalman temporal and measurement updates is applied to them, for they both map to an underlying sequence of Y estimates. The case we want to consider is when $(\hat{X}(t), P(t))$ is the result of a noiseless measurement update. Let $P'(t)$ be the x-reduced version of $P(t)$. Because the x rows and x columns of $P(t)$ are equal, we find that

$$P'(t) = MP'(t)M^{\mathrm{T}} = MP(t)M^{\mathrm{T}}.$$

Therefore, $(\hat{X}(t), P(t))$ and $(\hat{X}(t), P'(t))$ are consistent. Because consistency is an equivalence relation, it follows that x-reduction of the covariance matrix at date t and any subsequent dates never affects the future frequency and drift state estimates.

4.2. Reduced Kalman scale

X-reduction does affect the phase estimates, however. Just to check it out, the author tried it in a simulation, performing the x-reduction at each date. To the author's surprise, the Kalman scale defined by (10) no longer resembled the Kraw scale; it became as stable as the KPW scale. This procedure produces a new timescale, which we call the reduced Kalman scale, or Kred. To repeat, one simply carries out the Kalman filter, performing the x-reduction after each measurement update, and uses the common phase error (10) as the timescale Kred(t). Figure 1 shows the measured instability of Kred for our simulation example. In this view, Kred is hardly distinguishable from KPW; in fact, most of the Kred deviations are 1%–3% below the KPW deviations.

In the next two sections, we investigate what is behind this surprising empirical result.

4.3. The implicit weights

Weiss *et al* [4] showed that the Kraw scale is actually a weighted-average scale, with weights determined implicitly by the Kalman gain matrix. This is also true for the Kred scale, but the latter scale has different Kalman gains. Let us see how this works. The first (x_1) component of the Kalman measurement update equation can be written as

$$\hat{x}_1(t) = \tilde{x}_1(t) + \sum_{i=2}^{n} K_{1i}(t)[x_i(t) - x_1(t) - \tilde{x}_i(t) + \tilde{x}_1(t)], \tag{17}$$

where $K_{1i}(t)$ is the first row of the Kalman gain matrix, whose columns are indexed by the measurements $x_{i1}(t)$ for $i = 2$ to n. The $\tilde{x}_i(t)$ are the predicted phase estimates produced by the temporal update,

$$\tilde{x}_i(t) = \hat{x}_i(t-\delta) + \delta\hat{y}_i(t-\delta) + \tfrac{1}{2}\delta^2\hat{z}_i(t-\delta). \tag{18}$$

The implicit weights are defined by

$$\lambda_1(t) = 1+\sum_{i=2}^{n} K_{1i}(t), \qquad \lambda_i(t) = -K_{1i}(t) \quad \text{for } i = 2 \text{ to } n. \tag{19}$$

These weights add to 1, and (17) can be rewritten as

$$\hat{x}_1(t) = x_1(t) + \sum_{i=1}^{n} \lambda_i(t)[\tilde{x}_i(t) - x_i(t)]. \tag{20}$$

Finally, in (20) and (18) we may set

$$\hat{x}_1(t) = x_1(t) - x_e(t), \qquad \hat{x}_i(t-\delta) = x_i(t-\delta) - x_e(t-\delta),$$

where $x_e(t)$ is the Kalman scale defined by (10). When we do so, we find that (20) becomes the BTSE (11).

For our simulation example, it is interesting to show the explicit KPW weights, inversely proportional to r_i from (13), and the implicit weights of the Kraw and Kred scales, obtained by (19) from the Kalman gains at the end of the simulation:

	λ_{odd}	λ_{even}
Kraw	−0.0937	0.3437
KPW	0.2474	0.0026
Kred	0.2330	0.0170

The weights for the Kraw scale are inappropriate, to say the least; it even assigns negative weights to the odd-numbered clocks. The weights for the Kred scale are slightly less unbalanced than the KPW weights. Other simulations gave similar results. These observations led to the conjecture that the Kred weights are optimal in some sense. The next section shows that these weights indeed satisfy a precise optimality condition.

4.4. Optimality of the Kred scale

Returning to the expression (12) for the increment of a Kalman-based weighted-average timescale, we write it in the form

$$\Delta_\delta x_e(t) = \sum_{i=1}^{n} \lambda_i(t) N_i(t). \tag{21}$$

Let Γ be the covariance matrix of the random variables $N_i(t)$. As an aid to expressing it, write $p(y_i, y_j)$, $p(z_i, z_j)$, $p(y_i, z_j)$, $p(z_i, y_j)$ for the corresponding entries of the error covariance matrix $P(t-\delta)$ coming from the Kalman filter. The process–noises $w_{xi}(t)$, because they depend only on the underlying white noises between $t-\delta$ and t, are independent of the other random variables making up $N_i(t)$. Therefore,

$$\Gamma_{ij} = p(y_i, y_j)\delta^2 + [p(y_i, z_j) + p(z_i, y_j)]\frac{\delta^3}{2} + p(z_i, z_j)\frac{\delta^4}{4} + r_i\delta_{ij}, \tag{22}$$

where r_i is given by (13) and δ_{ij} is the Kronecker delta. This matrix is positive definite if all the r_i are positive, that is, if each clock has some noise.

Instead of just making the $\lambda_i(t)$ inversely proportional to r_i, as we do for the KPW scale, it is reasonable to choose them to minimize the variance of the scale increment $\Delta_\delta x_e(t)$. Letting λ be the row vector of the $\lambda_i(t)$, we want to minimize $\operatorname{var}\Delta_\delta x_e(t) = \lambda\Gamma\lambda^T$, given that $\sum\lambda_i = 1$. By using a Lagrange multiplier, or otherwise, we find that the optimal λ is such that the vector $\lambda\Gamma$ has entries that are all equal. Defining the $(n-1)\times n$ matrix

$$A = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \cdots & 0 \\ & & \cdots & & \ddots \\ -1 & 0 & 0 & \cdots & 1 \end{bmatrix},$$

we may assert that the optimal λ is the unique vector that satisfies

$$\lambda\Gamma A^T = 0, \qquad \sum\lambda_i = 1. \tag{23}$$

With this preparation, we can state and prove our main result, which goes a long way towards explaining the good short-term stability of the Kred scale.

Theorem 2. *Among all weighted-average timescales based on the Kalman frequency and drift state estimates, the Kred scale has the implicit weights that minimize the variance of the scale increment.*

This means that we do not have to solve (23) to obtain the optimal weights; they are delivered automatically by the x-reduced Kalman filter.

Proof. Define the auxiliary matrices

$$e_1 = [1\ 0\ \cdots\ 0]\ (1\times n),$$

$$B = \begin{bmatrix} 1\ 0\ 0 & & & \\ & 1\ 0\ 0 & & \\ & & \ddots & \\ & & & 1\ 0\ 0 \end{bmatrix} \quad (n\times(3n)).$$

Then $H = AB$. Let K_1 be the first row of the Kalman gain K. Let $\tilde{P} = \tilde{P}(t)$, the error covariance matrix after the temporal update from $t - \delta$ to t, and denote its first row by $\tilde{P}_1$. The Kalman gain satisfies the equation

$$KH\tilde{P}H^{\mathrm{T}} = \tilde{P}H^{\mathrm{T}}.$$

Therefore,

$$K_1 AB\tilde{P}B^{\mathrm{T}}A^{\mathrm{T}} = \tilde{P}_1 B^{\mathrm{T}}A^{\mathrm{T}}.$$

By the definition (19) of the implicit weights λ,

$$K_1 A = [1 - \lambda_1, -\lambda_2, \ldots, -\lambda_n] = e_1 - \lambda.$$

Also, $\tilde{P}_1 = e_1 B\tilde{P}$. Therefore,

$$(e_1 - \lambda)B\tilde{P}B^{\mathrm{T}}A^{\mathrm{T}} = e_1 B\tilde{P}B^{\mathrm{T}}A^{\mathrm{T}},$$

$$\lambda B\tilde{P}B^{\mathrm{T}}A^{\mathrm{T}} = 0.$$

We already know that $\sum \lambda_i = 1$. According to (23), we shall be done if we can prove that

$$B\tilde{P}B^{\mathrm{T}} = \Gamma. \tag{24}$$

From the temporal update equation for the covariance, we have

$$\tilde{P} = \Phi P(t - \delta)\Phi^{\mathrm{T}} + Q,$$

$$B\tilde{P}B^{\mathrm{T}} = B\Phi P(t - \delta)\Phi^{\mathrm{T}}B^{\mathrm{T}} + BQB^{\mathrm{T}}.$$

Now

$$B\Phi = \begin{bmatrix} 1 & \delta & \frac{\delta^2}{2} & & & & & & \\ & & & 1 & \delta & \frac{\delta^2}{2} & & & \\ & & & & & & \ddots & & \\ & & & & & & 1 & \delta & \frac{\delta^2}{2} \end{bmatrix}.$$

Because $P(t - \delta)$ is x-reduced, it is composed of the 3×3 submatrices

$$P_{ij} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & p(y_i, y_j) & p(y_i, z_j) \\ 0 & p(z_i, y_j) & p(z_i, z_j) \end{bmatrix}.$$

The (x_i, x_j) element of Q is $r_i \delta_{ij}$. Thus, the (i, j) element of $B\tilde{P}B^{\mathrm{T}}$ is just

$$\begin{bmatrix} 1 & \delta & \frac{\delta^2}{2} \end{bmatrix} P_{ij} \begin{bmatrix} 1 & \delta & \frac{\delta^2}{2} \end{bmatrix}^{\mathrm{T}} + (BQB^{\mathrm{T}})_{ij} = \Gamma_{ij}$$

from (22). This establishes (24), and the proof is complete. ∎

5. Final remarks

We have discussed the properties of three timescales based on the Jones–Tryon Kalman filter. All three scales are actually weighted-average scales; they use the same Kalman frequency and drift state estimates, but have different weights. The Kraw scale, which was used for TA(NIST), is just the common phase error of the unmodified Kalman filter. The KPW scale ignores the Kalman phase estimates entirely and uses the frequency and drift state estimates explicitly in the BTSE, with weights determined by an auxiliary formula. We have seen, though, that the KPW scale can be regarded as an approximation to the Kred scale, in which we reach inside the Kalman filter to reduce the covariance matrix at each step, then use the phase estimates in the same way that they are used for the Kraw scale. The implicit weights of the resulting scale are optimal in a sense just described. Both the KPW and Kred scales perform well over the whole range of averaging times. The slight advantage of the Kraw scale in a range of long averaging times is not a good trade-off for its short-term instability.

This study has been carried out in a simulation playpen in which all clocks behave according to their assumed models. A practical timescale must deal with outliers, jumps, and ensemble changes, and it ought to provide for adaptive estimation of the qs. One might also want to add white PM and measurement noise to the model. Management of a Kalman-based timescale presents practical difficulties, especially the problem of inserting a new clock into the ensemble without causing a phase or frequency step in the scale. Perhaps the simplicity and excellent stability of the Kred scale, under ideal conditions to be sure, will motivate an effort to overcome these difficulties.

Acknowledgments

Thanks to P Tavella for her advice and encouragement. This work was performed by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

[1] Barnes J A, Jones R H, Tryon P V and Allan D W 1982 Stochastic models for atomic clocks *Proc. 14th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting* pp 295–306
[2] Jones R H and Tryon P V 1987 Continuous time series models for unequally spaced data applied to modeling atomic clocks *SIAM J. Sci. Stat. Comput.* **8** 71–81
[3] Brown K R Jr 1991 The theory of the GPS composite clock *Proc. ION GPS-91 Meeting* pp 223–41
[4] Weiss M A, Allan D W and Peppler T K 1989 A study of the NBS time scale algorithm *IEEE Trans. Instrum. Meas.* **38** 631–5
[5] Weiss M and Weissert T 1991 AT2, a new time scale algorithm: AT1 plus frequency variance *Metrologia* **28** 65–74
[6] Hutsell S T 1995 Relating the Hadamard variance to MCS Kalman filter clock estimation *Proc. 27th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting* pp 291–301
[7] Gelb A (ed) 1974 *Applied Optimal Estimation* (Cambridge MA: MIT Press)
[8] Bierman G J 1977 *Factorization Methods for Discrete Sequential Estimation* (New York: Academic)
[9] Grewal M S and Andrews A P 2001 *Kalman Filtering: Theory and Practice Using Matlab* (New York: Wiley)
[10] *Guide to the Expression of Uncertainty in Measurement* 1995 (Geneva: International Organization for Standardization)
[11] Greenhall C A 2001 Kalman plus weights: a time scale algorithm *Proc. 33rd Annual Precise Time and Time Interval (PTTI) Systems and Applications Meeting* pp 445–54

[12] Stein S R 1992 Advances in time-scale algorithms *Proc. 24th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting* pp 289–302

[13] *Selection and Use of Precise Frequency and Time Systems* 1997 (Geneva: International Telecommunication Union)

[14] Breakiron L A 1991 Timescale algorithms combining cesium clocks and hydrogen masers *Proc. 23rd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting* pp 297–305

[15] Guinot B 1987 Some properties of algorithms for atomic time scales *Metrologia* **24** 195–8

[16] Satin A L, Feess W A, Fliegel H F and Yinger C H 1990 GPS composite clock software performance *Proc. 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting* pp 529–45