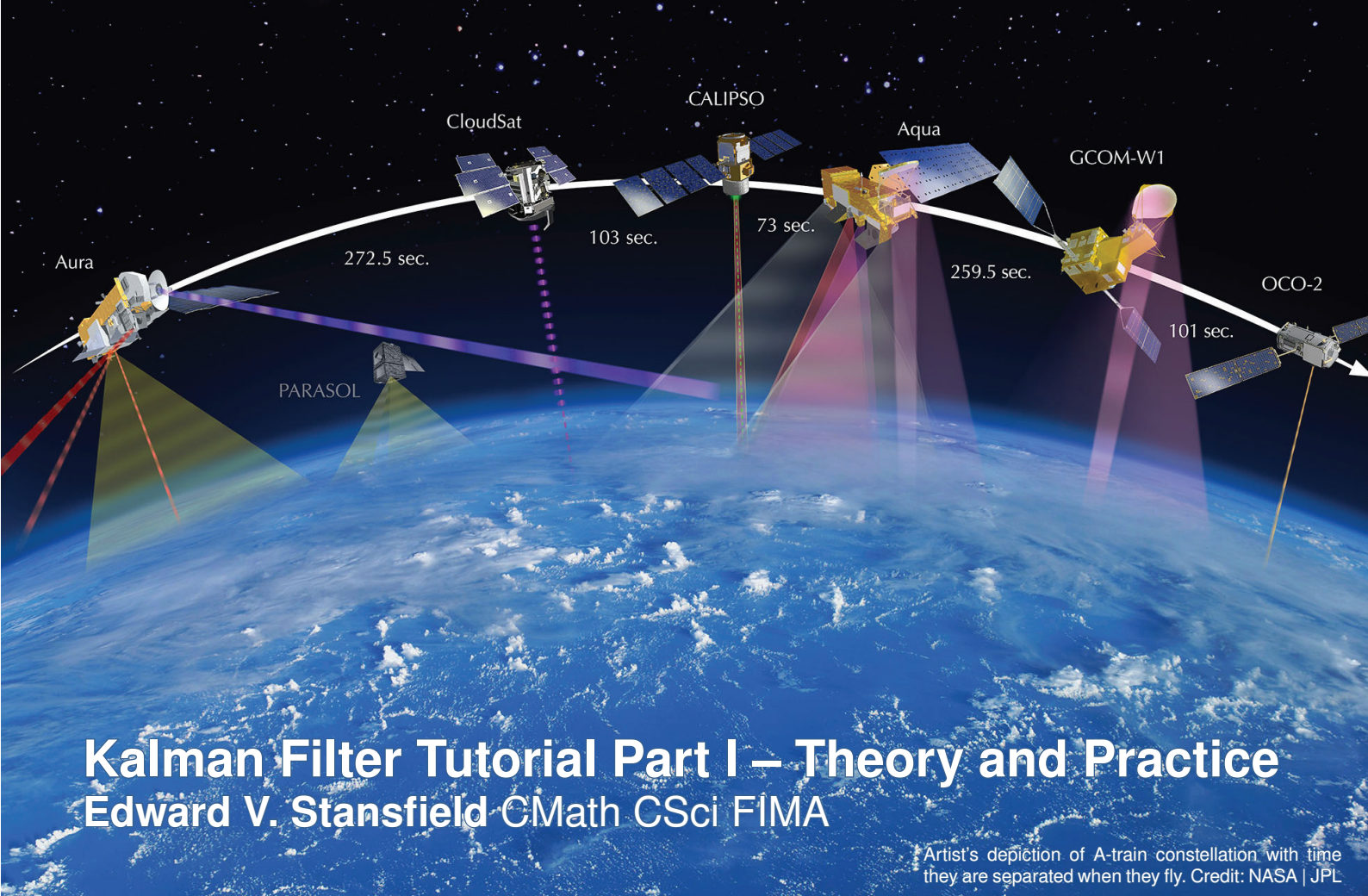




Kalman Filter Tutorial Part I – Theory and Practice

Edward V. Stansfield CMath CSci FIMA

Artist's depiction of A-train constellation with time they are separated when they fly. Credit: NASA | JPL

Introduction

The purpose of this article is to provide an introduction to Kalman filters [1, 2], focusing on what they are and what they can do. The intention is to keep the mathematical discussion as simple as possible whilst retaining sufficient rigour to illustrate many interesting and revealing relationships between the various quantities that can be computed. The article is based on one written in March 2001 for a symposium organised by the author and a colleague. It was published on a website, but was not archived and is no longer available. The opportunity has been taken here to restructure the contents and outline some developments that have occurred since then.

The term Kalman filter is traditional, but it is not a filter in the conventional sense, since it does not, for example, modify a signal to reduce noise or eliminate unwanted components. A Kalman filter is based on a mathematical model of a dynamic system, and uses measurements on the system to provide information about how it behaves over time. Depending on how the model is used, a Kalman filter may either smooth previous behaviour, or estimate current behaviour, or predict future behaviour. In Kalman filter literature it is common to refer to the dynamic system as the *plant* and to the measurements on the system as *observations*. In this article, the terms *dynamic system* and *plant* have the same meaning and are used interchangeably, as are the terms *measurement* and *observation*.

A common application of Kalman filters is to track moving targets, in which case the dynamic system is the actual target track. If the measurements are made in Cartesian coordinates,

and if the measurement errors can be regarded as independent and normally distributed (i.e. Gaussian), and if the target behaviour is known a priori, then a Kalman filter will provide a minimum mean squared error (MMSE) estimate of target position and velocity. If any of these prior conditions are invalid, then a form of Kalman filter can still be used, but the results may be sub-optimum in the sense that the estimated position and velocity will not necessarily meet the MMSE criterion.

The basic Kalman filter is first introduced, with all detailed mathematical derivations provided in appendixes (available online at www.ima.org.uk/viewitem.cfm-cit_id=384305.html). A simple target tracking example is used to illustrate the process and the resulting performance when both the measurements and the system contain additive noise. In Part II (August 2016) some variants and extensions of the basic Kalman filter are described in outline to give a flavour of what they offer over and above the basic Kalman filter. However, in keeping with the introductory tutorial intentions of this article, detailed mathematical analysis of these extensions is not provided. Finally, some of the many applications for Kalman filters are briefly described.

... a Kalman filter may either smooth previous behaviour, or estimate current behaviour, or predict future behaviour ...

Kalman filter basics

Plant model

A simple Kalman filter target tracker is based on the assumption that the motion can be modelled as a point target moving in a straight line with constant velocity. In reality, the target velocity and direction will not be constants, and in order to allow for this a noise component (and any known forcing function)

is included in the model. If we denote the target position and velocity at time t by the state vector $\mathbf{x}(t)$, then under the constant velocity assumption the state at time $t + T$ will be given by $\mathbf{x}(t + T) = \Phi(T)\mathbf{x}(t) + \mathbf{w}(t)$ where $\Phi(T)$ is the *state transition matrix* for the time interval T , and $\mathbf{w}(t)$ is the additive noise component, assumed to be Gaussian. Note that for simplicity this equation does not include any forcing function. In detail, if we consider the target moving in a plane with Cartesian coordinates (x_1, x_2) , then in matrix form the mathematical model for target motion might be given by

Target motion:

$$\mathbf{x}(t + T) = \Phi(T)\mathbf{x}(t) + \mathbf{w}(t) \Rightarrow \begin{pmatrix} x_1(t + T) \\ x_2(t + T) \\ \dot{x}_1(t + T) \\ \dot{x}_2(t + T) \end{pmatrix} = \begin{pmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \\ \dot{x}_1(t) \\ \dot{x}_2(t) \end{pmatrix} + \begin{pmatrix} w_1(t) \\ w_2(t) \\ \dot{w}_1(t) \\ \dot{w}_2(t) \end{pmatrix}.$$

The plant (target model) noise covariance matrix $\mathbf{Q}(t) = E[\mathbf{w}(t)\mathbf{w}(t)^T]$ is assumed to be known, where superscript T denotes a matrix transpose and $E[\]$ denotes statistical expectation.

Observation model

If the measurement vector on the target is denoted by $\mathbf{y}(t)$, this is assumed to be related to the actual state vector $\mathbf{x}(t)$ via the measurement equation $\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t) + \mathbf{v}(t)$ where $\mathbf{H}$ is the measurement matrix and $\mathbf{v}(t)$ is the measurement noise, assumed to be independent of the state noise $\mathbf{w}(t)$ and Gaussian. In the simple case of motion in a plane with constant velocity, then $\mathbf{H}$ will be an identity matrix of dimension four if both position and velocity are measured directly. If position or velocity, but not both, is measured, then $\mathbf{H}$ takes the form of a 2×4 matrix containing an identity matrix of dimension two as a sub-matrix. For example, if only position is measured, then in detail the measurement equation will be

Position measurement:

$$\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t) + \mathbf{v}(t) \Rightarrow \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \\ \dot{x}_1(t) \\ \dot{x}_2(t) \end{pmatrix} + \begin{pmatrix} v_1(t) \\ v_2(t) \end{pmatrix}.$$

The measurement noise $\mathbf{v}(t)$ is assumed to have a known covariance $\mathbf{R}(t) = E[\mathbf{v}(t)\mathbf{v}(t)^T]$. Mathematical models of the target motion and measurements for the simple Kalman filter can thus be summarised in the following two equations:

State (plant) equation:

$$\mathbf{x}(t + T) = \Phi(T)\mathbf{x}(t) + \mathbf{w}(t), \quad \mathbf{Q}(t) = E[\mathbf{w}(t)\mathbf{w}(t)^T],$$

Measurement (observation) equation:

$$\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t) + \mathbf{v}(t), \quad \mathbf{R}(t) = E[\mathbf{v}(t)\mathbf{v}(t)^T].$$

Notation

The notation traditionally used for Kalman filters is at first sight cumbersome, but experience shows that it generally provides the greatest insight into the process of forming recursive least squares estimates. It is as follows: the value of a vector $\mathbf{x}(t)$ at a discrete time instant $t = t_k$ is denoted by $\mathbf{x}_k$. A best estimate of $\mathbf{x}(t)$ at time $t = t_k$ based on observations up to a time $t = t_j$ is denoted by $\mathbf{x}_{k|j}$.

This notation encompasses the three cases of smoothing ($k < j$), filtering ($k = j$), and prediction ($k > j$). In many applications only estimation and prediction¹ are involved, but there are occasions when smoothing could also be useful. For example, it may be required to find the position that a satellite occupied at some time in the past. In such cases, observations made both before and after the required time are available, and smoothing will be an appropriate estimation process.

In applications where smoothing is not required, an often used simpler notation is as follows: when $j = k$ the *best estimate* is $\mathbf{x}_{k|k} = \hat{\mathbf{x}}_k$, and when $j = k - 1$ the *best one-step prediction* is $\mathbf{x}_{k|k-1} = \tilde{\mathbf{x}}_k$.

Estimator

Kalman filter processing has as its objective estimation of the state vector $\mathbf{x}(t_k)$ at time $t = t_k$ based on measurements $\mathbf{y}(t)$ taken at times $t = t_n$ for $n = 0, 1, \dots, k$. It is a sequential process in the sense that each new measurement updates a previous prediction $\tilde{\mathbf{x}}(t_k)$ by processing the new information $\mathbf{y}(t_k)$ to form a new estimate $\hat{\mathbf{x}}(t_k)$. This is accomplished by forming an innovation $\mathbf{z}(t_k)$ as the difference between the actual measurement $\mathbf{y}(t_k)$ and a prediction of it based on the previous measurements. The terminology here stems from the fact that the innovation represents the new information (i.e. that which was not predicted) contained in the measurement.

The actual equations for a simple Kalman filter (one which does not involve smoothing) can be written in the form

$$\begin{aligned} \text{Innovation:} \quad & \mathbf{z}_k = \mathbf{y}_k - \mathbf{H}_k \tilde{\mathbf{x}}_k, \\ \text{Estimate:} \quad & \hat{\mathbf{x}}_k = \tilde{\mathbf{x}}_k + \mathbf{K}_k \mathbf{z}_k, \\ \text{Prediction:} \quad & \tilde{\mathbf{x}}_{k+1} = \Phi_k \hat{\mathbf{x}}_k, \end{aligned}$$

$$\begin{aligned} \text{Kalman gain:} \quad & \mathbf{K}_k = \tilde{\mathbf{P}}_k \mathbf{H}_k^T [\mathbf{H}_k \tilde{\mathbf{P}}_k \mathbf{H}_k^T + \mathbf{R}_k]^{-1}, \\ \text{Estimate covariance:} \quad & \hat{\mathbf{P}}_k = [\mathbf{I} - \mathbf{K}_k \mathbf{H}_k] \tilde{\mathbf{P}}_k, \\ \text{Prediction covariance:} \quad & \tilde{\mathbf{P}}_{k+1} = \Phi_k \hat{\mathbf{P}}_k \Phi_k^T + \mathbf{Q}_k. \end{aligned}$$

A derivation from first principles of these equations is provided in Appendix 1. Figure 1 is a block diagram which shows both the underlying models and the Kalman filter processing required to produce an optimum estimate of the state from the observations. Note that the diagram also shows how a known forcing function $\mathbf{f}_k$ can be readily taken into account by the state estimator.

The equations have been expressed in their most general form, in the sense that the state transition matrix Φ and the measurement matrix $\mathbf{H}$ are not constrained to be constant. They are also

... these covariances can be used to control the rate of convergence of the Kalman filter estimator, with the optimum values depending on the manoeuvrability of the target being tracked ...

general in that no limitations have been imposed on the dimensionality of the matrices. This is of course subject to the primary conditions of the model on which they are based, namely that the state and measurements equations are linear, and that the state and measurement noises are independent with Gaussian distributions.

It will be observed from the above that the lower equations for the Kalman filter (Kalman gain, estimate and prediction covariances) are independent of the upper ones (innovation, estimate and prediction). This indicates that the estimated error covariance $\hat{\mathbf{P}}$ of the state estimate $\hat{\mathbf{x}}$ is independent of the measurements. Indeed, if desired, the error covariance could even be pre-computed before any measurements are made. This is a direct consequence of the underlying assumption that the noise covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ are assumed to be known a priori. In practice, these covariances can be used to control the rate of convergence of the Kalman filter estimator, with the optimum values depending on the manoeuvrability of the target being tracked.

Exemplar – one-dimensional motion

By way of illustration, an example of a Kalman filter used to track a moving particle is described for the case of one-dimensional linear motion under the influence of acceleration excitation. Although this is arguably the simplest example that can be chosen, it nonetheless encapsulates the basic principles of a Kalman filter and illustrates many of its important properties.

Solution of linear differential equation

Consider the simple dynamic model of particle motion obeying the constant-coefficient first-order linear differential equation

$$\dot{\mathbf{X}}(t) = \mathbf{F}\mathbf{X}(t) + \mathbf{G}(t)$$

where

$$\mathbf{X}(t) = \begin{pmatrix} x(t) \\ \dot{x}(t) \\ \ddot{x}(t) \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{G}(t) = \begin{pmatrix} 0 \\ 0 \\ g(t) \end{pmatrix}.$$

The scalar quantity $x(t)$ is the position of the particle at time t , and $g(t)$ is the manoeuvre excitation. If $g(t) = 0$ then $x(t)$ describes the unconstrained motion of a particle moving with constant velocity. The precise value of the velocity is determined by the initial condition of the motion. Changes in the velocity occur whenever $g(t)$ is not zero.

A general solution of this first-order vector linear differential equation is given by

$$\mathbf{X}(t) = \Phi(t, t-T)\mathbf{X}(t-T) + \mathbf{W}(t, T)$$

where

$$\mathbf{W}(t, T) = \int_{t-T}^t \Phi(t, \lambda)\mathbf{G}(\lambda)d\lambda.$$

The form of the solution imposes certain conditions on the transition matrix $\Phi(t, \lambda)$. These are that it is differentiable, and satisfies the following three equations:

$$\begin{aligned} \Phi(t, t) &= \mathbf{I}, \quad \Phi(t, \lambda) = \Phi(t, \mu)\Phi(\mu, \lambda), \\ \frac{\partial \Phi(t, \lambda)}{\partial t} &= \mathbf{F}\Phi(t, \lambda). \end{aligned}$$

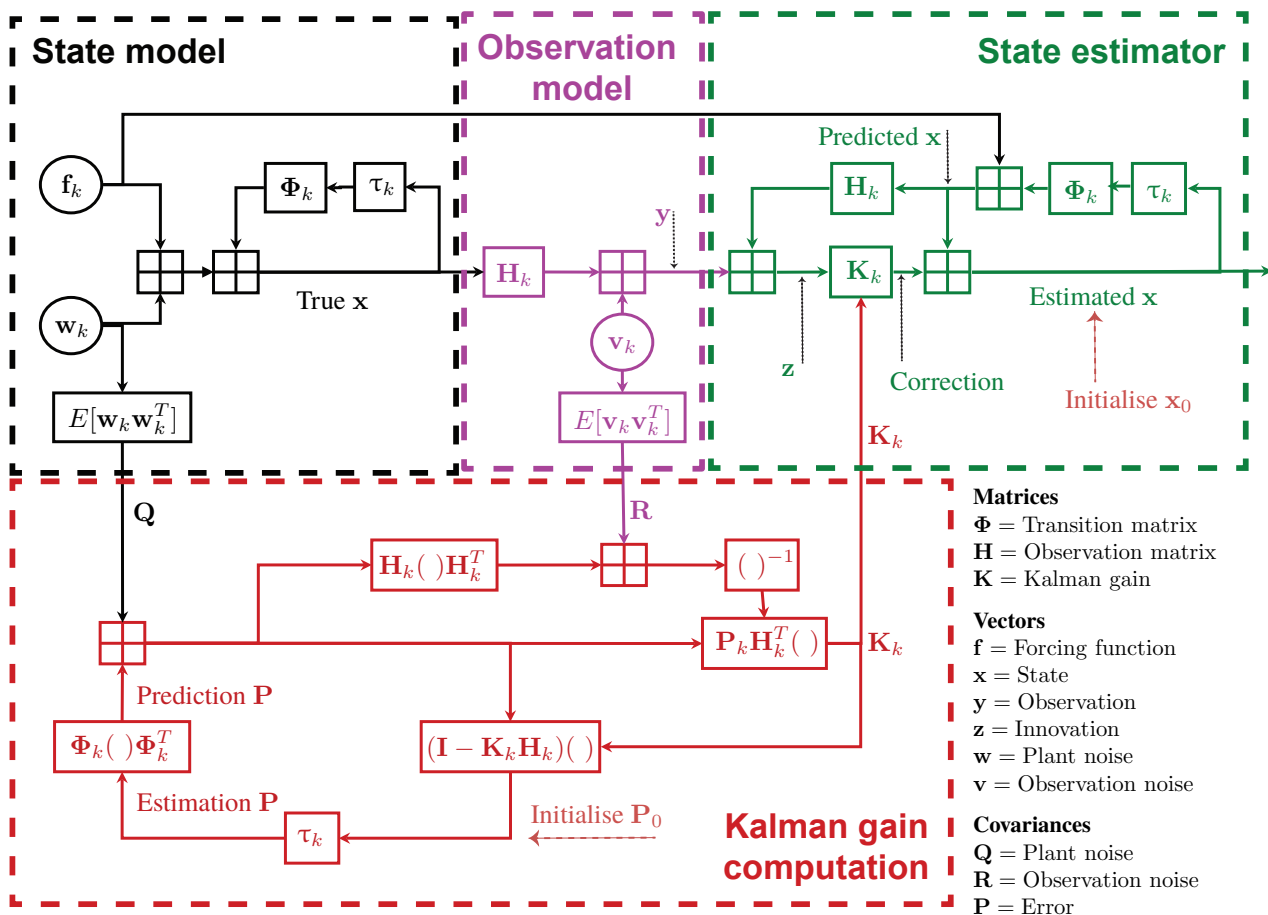


Figure 1: Kalman filter process model.

The first equation follows immediately from the form of the solution when $T = 0$. The second equation is a continuity condition which ensures the validity of the solution at all times intermediate between $t - T$ and t . The third equation defines the transition matrix $\Phi(t, \lambda)$ as the solution of a first-order linear differential equation, and is given by

$$\Phi(t, \lambda) = \exp[(t - \lambda)\mathbf{F}] = \mathbf{I} + (t - \lambda)\mathbf{F} + \frac{(t - \lambda)^2\mathbf{F}^2}{2!} + \frac{(t - \lambda)^3\mathbf{F}^3}{3!} + \frac{(t - \lambda)^4\mathbf{F}^4}{4!} + \dots$$

It is easy to show that

$$\mathbf{F} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{F}^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{F}^n = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \forall n \geq 3$$

so that

$$\Phi(t, \lambda) = \mathbf{I} + (t - \lambda)\mathbf{F} + \frac{(t - \lambda)^2}{2}\mathbf{F}^2$$

$$= \begin{pmatrix} 1 & t - \lambda & \frac{1}{2}(t - \lambda)^2 \\ 0 & 1 & t - \lambda \\ 0 & 0 & 1 \end{pmatrix}.$$

With $\lambda = t - T$, the transition matrix depends solely on the time lag T , and hence one can write

$$\mathbf{X}(t) = \Phi(T)\mathbf{X}(t - T) + \mathbf{W}(t, T)$$

where

$$\Phi(T) = \mathbf{I} + T\mathbf{F} + \frac{T^2}{2}\mathbf{F}^2 = \begin{pmatrix} 1 & T & \frac{1}{2}T^2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{pmatrix}.$$

Note that with this formulation of the transition matrix, the plant noise component can be expressed in the form

$$\mathbf{W}(t, T) = \int_{t-T}^t \Phi(t - \lambda)\mathbf{G}(\lambda)d\lambda.$$

Noise covariance

When a Kalman filter is used to track the motion of a particle based on measurements of its position and/or velocity, the applied acceleration excitation may not be known. Under these circumstances, one often used technique is to model the excitation $g(t)$ as a white stochastic random process with zero mean and a Gaussian distribution. In the absence of any information on the manoeuvrability of the particle, the variance of this process can only be guessed at. In what follows, it will be assumed that the continuous random process $g(t)$ has an unknown standard deviation σ , and that its covariance can be modelled as:

$$E[g(\lambda)g(\mu)] = \sigma^2\delta(\lambda - \mu)$$

where $\delta(\cdot)$ is the Dirac delta function. It remains now to determine the covariance $\mathbf{Q}$ of the discrete time noise process $\mathbf{W}(t, T)$. This is defined by

$$\mathbf{Q} = E[\mathbf{W}(t, T)\mathbf{W}(t, T)^T]$$

$$= E\left[\int_{\lambda=t-T}^t \int_{\mu=t-T}^t \Phi(t - \lambda)\mathbf{G}(\lambda)\mathbf{G}(\mu)^T\Phi(t - \mu)^T d\lambda d\mu\right].$$

Since the expectation operator is linear, it will operate only on the factor $\mathbf{G}(\lambda)\mathbf{G}(\lambda)^T$, and hence we can write

$$\mathbf{Q} = \sigma^2 \int_{\lambda=t-T}^t \Phi(t - \lambda) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \Phi(t - \lambda)^T d\lambda.$$

Setting $t - \lambda = \gamma$, changing the integration variable to γ , expanding and integrating then obtains

$$\mathbf{Q} = \sigma^2 T \begin{pmatrix} \frac{1}{20}T^4 & \frac{1}{8}T^3 & \frac{1}{6}T^2 \\ \frac{1}{8}T^3 & \frac{1}{3}T^2 & \frac{1}{2}T \\ \frac{1}{6}T^2 & \frac{1}{2}T & 1 \end{pmatrix}.$$

This form of noise covariance is often used when Kalman filters are applied to the problem of tracking a moving target whose acceleration excitation is unknown. If the target is known to be highly manoeuvrable, then a value of σ corresponding to its expected high acceleration is appropriate, but for targets whose manoeuvrability is low a value of σ corresponding to its expected low acceleration is appropriate.

Simulation

In order to illustrate the effect of using different values of standard deviation σ for target acceleration, a computer simulation was undertaken to track a target subjected to a known acceleration. Specifically, a one-dimensional target motion was generated for the following circumstances:

- Stationary for 5 seconds from $t = 0$ to $t = 5$ seconds.
- Constant acceleration of 10 m/s^2 for 10 seconds from $t = 5$ to $t = 15$ seconds.
- Constant velocity for a further 15 seconds from $t = 15$ to $t = 30$ seconds.

Simulated measurements of both the position and velocity of the target were generated at 10-millisecond intervals i.e. $T = 0.01$ seconds. Measurement errors were assumed to have a Gaussian distribution with zero mean, and be independent from sample to sample. The standard error of position measurements was taken to be 100 m, and that of velocity measurements was 4 m/s.

The measurements were fed to a Kalman filter, which then estimated the target motion from the simulated measurements. The filter state was initialised to the correct value $X_0 = (0, 0, 0)^T$, and the initial predicted covariance matrix was somewhat arbitrarily set to be

$$\tilde{\mathbf{P}}_0 = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.1 \end{pmatrix}.$$

The results of the simulation are shown in Figure 2 using two different values for the standard deviation σ of the target acceleration.

The left-hand column shows the simulated position, velocity and acceleration of the target. The middle column shows the error in the Kalman filter estimate of these quantities when the standard deviation of the target acceleration is assumed to be

$\sigma = 0.003 \text{ m/s}^2$. The right-hand column shows the same quantities if a value of $\sigma = 3 \text{ m/s}^2$ is used instead. It is readily seen that the smaller value of σ results in larger estimation errors compared with the larger value of σ . Moreover, the errors are smoother for the smaller value of σ than they are for the larger value. In a sense, the choice of σ can be thought of as being analogous to setting the cut-off frequency of a low-pass filter acting on the estimation error.

The bottom two graphs of Figure 2 show the uniformly most powerful invariant test statistics (UMPITS) computed from the Kalman filter innovation sequences (see Appendix 3 for details). The magnitude of these statistics clearly indicates that the smaller value of target acceleration standard deviation σ results in larger errors than the larger value of σ . The statistic is thus shown to be a useful indicator of Kalman filter performance.

One final comment: although the Kalman filter model does not match the reality that the target has a deterministic acceleration, and is thus not random, the results nonetheless show a tracking error that is significantly less than the measurement error. Note, however, that even when the filter is performing reasonably

well (i.e. when the assumed target acceleration is $\sigma = 3 \text{ m/s}^2$, shown in the right-hand column of Figure 2), the actual value of the UMPITS far exceeds its theoretical value, which is the dimension of the measurement vector (2 in this case). This is thought to be a direct consequence of the Kalman filter being based on a model of target motion that is different from the actual target behaviour.

Notes

1. In this context a prediction is a forecast of a future value, whereas an estimate refers to a current value.

REFERENCES

- 1 Kalman, R.E. (1960) A new approach to linear filtering and prediction problems, *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45.
- 2 Brookner, E. (1998) *Tracking and Kalman Filtering Made Easy*, John Wiley & Sons Inc., New Jersey.

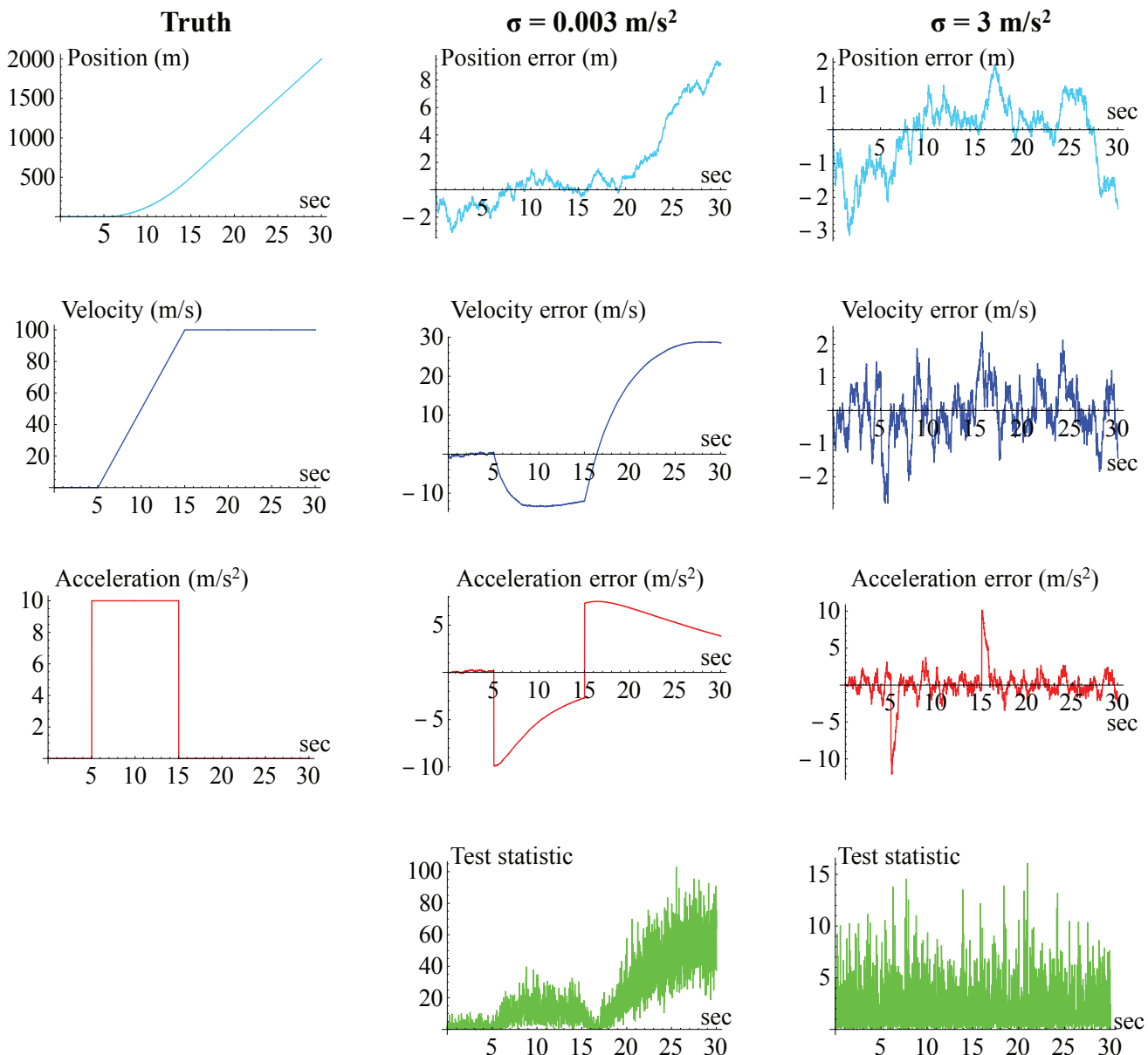


Figure 2: Tracking a target with two different standard deviations.