

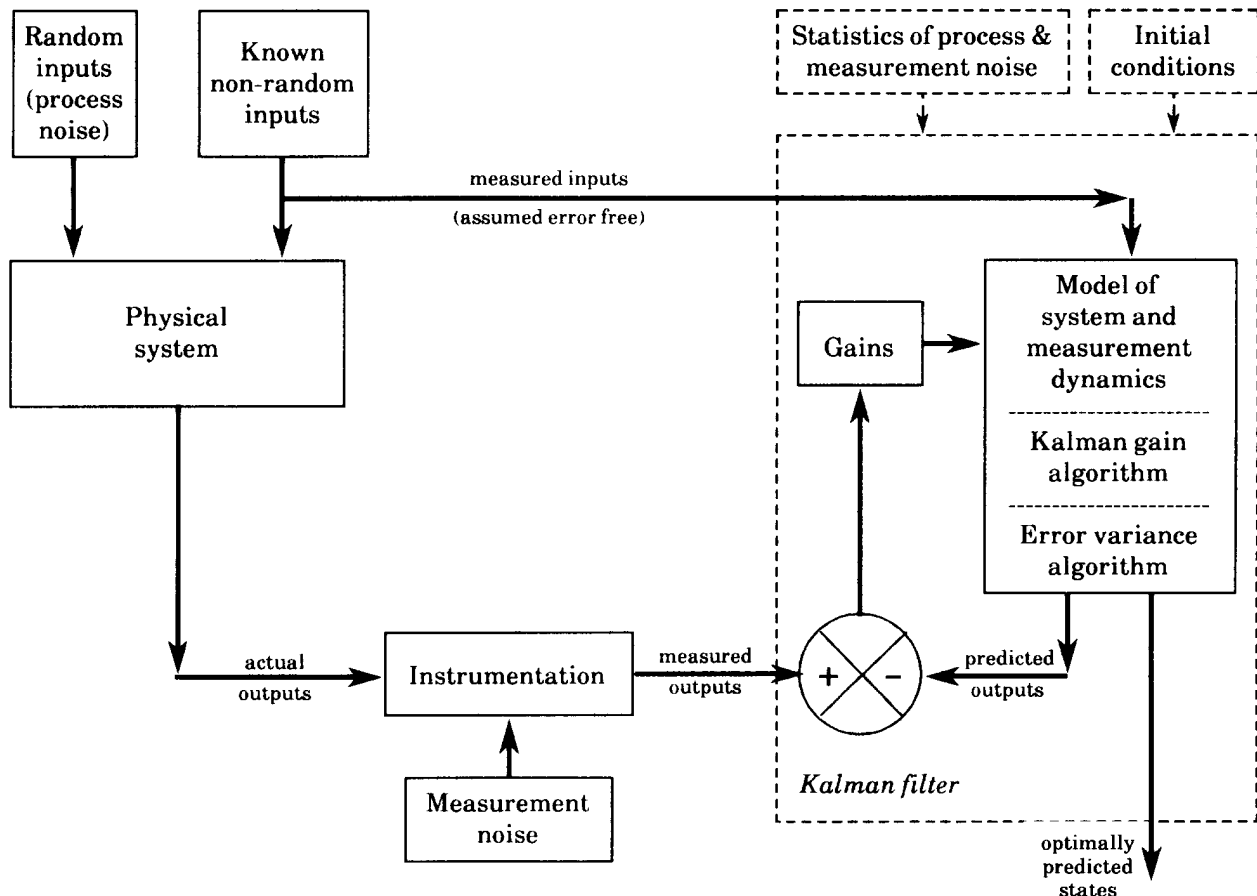
THE KALMAN FILTER

1. INTRODUCTION

This Data Item presents a derivation and description of the Kalman filter, one of the most common filtering techniques used for estimating the state(s) of a randomly excited linear system. The Kalman filter, termed an optimal estimator, is a computational algorithm that attempts to deduce the random time varying states of a system from the time varying measurements of those states by filtering out the measurement noise. The Kalman filter is, consequently, not a smoothing filter. The Kalman filter utilises knowledge of:

- (a) measurements of the state(s) of the system or a linear combination thereof (corrupted by measurement noise),
 - (b) model of the system (which incorporates random process noise),
 - (c) statistics of the system and measurement noise,
 - (d) initial conditions
- and (e) deterministic input(s).

A schematic illustration of the Kalman filter is shown in Sketch 1.1 in which the dotted box performs the function of the filter. The known inputs to the physical system are the same as those to the mathematical model of the system. Outputs from the model and the measurements are compared and the differences are multiplied by a set of gains that are fed back to the appropriate parts of the model. The statistics of the process noise and the measurement noise together with the initial conditions are required by the filter. The outputs of the filter are the optimally predicted system states.



Sketch 1.1 Illustration of Kalman filter

Issued December 1988

Optimal estimates produced by the Kalman filter are those that appropriately weight a combination of measurement and prediction in a manner that minimises the variance of the estimation error of the system states. The variances of the system noise (or process noise) and the measurement noise needs to be known a priori or are otherwise available. The Kalman filter technique is termed optimal since it provides an error variance that is less than or equal to that found by any other unbiased estimate.

The advantage of this kind of filter is its ability to define a method of calculating the gains such that the optimal estimates are obtained whilst maintaining a running account of the expected variances of the errors in these estimates. Corresponding disadvantages include the sensitivity of the filter to both erroneous a priori models and erroneous statistics of the process noise and the measurement noise.

This Item derives the discrete and continuous forms of the Kalman filter equations for a first-order randomly excited linear system. Equations for the multi-dimensional filter used for a randomly excited system comprising n states are also given.

An example of the pure rolling motion of an aircraft illustrates the application of the discrete filter.

2. NOTATION

Any coherent system of units may be used.

$\underline{A}$	plant or stability matrix associated with continuous form of state equation
a	coefficient of state variable, first-order continuous state equation
$\underline{B}$	control or input matrix associated with continuous form of state equation
b	coefficient of input variable, first-order continuous state equation
$\underline{C}$	matrix associated with continuous form of observation equation
c	coefficient of state variable, first-order continuous observation equation
$\underline{D}$	coefficient matrix of process noise associated with continuous form of state equation
$E\{\dots\}$	expected values of $\{\dots\}$
e	coefficient of process noise, first-order continuous form of state equation
h	general weighting parameter
$\underline{I}$	unit matrix
$\underline{K}$	matrix of Kalman gains
k	Kalman gain
l	number of measured variables
m	number of input variables
n	number of state variables

$\underline{P}$	covariance matrix of state variables
p	number of process noise sequences
$\underline{Q}$	covariance matrix of process
$\underline{R}$	covariance matrix of measurement noise sources
t	time
u	input variable
$V\{\dots\}$	variance of $\{\dots\}$
v	measurement noise sequence
w	process noise sequence
x	state variable
z	measured output quantity
β	estimation error
$\delta(t - \tau)$	Dirac delta function
ε	error quantity
$\underline{\Theta}$	matrix associated with discrete form of observation equation
θ	coefficient of state variable, first-order discrete observation equation
$\underline{\Lambda}$	coefficient matrix of process noise associated with discrete form of state equation
λ	coefficient of process noise, first-order discrete state equation
τ	sampling time, $t_{i+1} - t_i$
σ	standard deviation
$\underline{\Phi}$	plant or stability matrix associated with discrete form of state equation
ϕ	coefficient of state variable, first-order discrete state equation
$\underline{\Psi}$	control or input matrix associated with discrete form of state equation
ψ	coefficient of input variable, first-order discrete state equation

Subscripts

f	denotes forecast
i	denotes discrete time points
opt	optimal

v refers to measurement noise

w refers to process noise

Superscript

T denotes transform

A bar below a quantity denotes a matrix or vector quantity.

A bar above a quantity denotes a mean or average quantity.

3. OPTIMAL ESTIMATION

The Kalman filter is an optimal estimation technique that provides an unbiased, minimum variance estimate for each of the states of a linear system from a combination of measurements and model predictions.

3.1 System Model

Consider the following general equations for the model of a linear dynamical system

$$\dot{\underline{x}}(t) = \underline{A}(t) \underline{x}(t) + \underline{B}(t) \underline{u}(t) + \underline{D}(t) \underline{w}(t), \quad \underline{x}(t_0) = \underline{x}_0, \quad (3.1)$$

where $\underline{x}(t)$ is the state vector of size $(n \times 1)$, with elements given by

$$\underline{x}(t) = [x(1, t), x(2, t), \dots, x(n, t)]^T,$$

$\underline{u}(t)$ is the known (deterministic) input vector of size $(m \times 1)$, with elements given by

$$\underline{u}(t) = [u(1, t), u(2, t), \dots, u(m, t)]^T$$

and $\underline{w}(t)$ is the noise vector of size $(p \times 1)$ representing p independent random white noise inputs, with elements given by

$$\underline{w}(t) = [w(1, t), w(2, t), \dots, w(p, t)]^T.$$

The white noise inputs, referred to as process, or system, noise, have the following known statistics:

$$E\{\underline{w}(t)\} = 0, \text{ for all } t$$

and
$$E\{\underline{w}(\tau) \underline{w}(t)^T\} = \underline{Q}(t) \delta(t - \tau),$$

where $\delta(t - \tau)$ is the Dirac delta function. It is usual to call $\underline{A}$ which is of size $(n \times n)$ the plant or stability matrix, $\underline{B}$ which is of size $(n \times m)$ the input or control matrix and $\underline{D}$ which is of size $(n \times p)$ the system noise matrix.

Measurements, $\underline{z}(t)$, of the system are related to the system states by

$$\underline{z}(t) = \underline{C}(t) \underline{x}(t) + \underline{v}(t), \quad (3.2)$$

where $\underline{v}(t)$ is a vector of size $(l \times 1)$ representing l independent random white noise sources of measurement noise with known statistics,

$$E\{\underline{v}(t)\} = 0, \text{ for all } t$$

$$\text{and } E\{\underline{v}(\tau)\underline{v}(\tau)^T\} = \underline{R}\delta(t - \tau).$$

The process noise and the measurement noise are assumed to be independent white noise sources and so

$$E\{\underline{w}(t)\underline{v}(t')\} = 0, \text{ for all } t \text{ and } t'.$$

The discrete forms equivalent to Equations (3.1) and (3.2) are

$$\underline{x}_{i+1} = \underline{\Phi}_i \underline{x}_i + \underline{\Psi}_i \underline{u}_i + \underline{\Lambda}_i \underline{w}_i \quad (3.3)$$

$$\text{and } \underline{z}_{i+1} = \underline{\Theta}_i \underline{x}_{i+1} + \underline{v}_{i+1}, \quad (3.4)$$

where i is an integer denoting the individual value of t_i . The relationships between the coefficients of the continuous system and those of the equivalent discrete system, when it is assumed that the sampling interval, $\tau (\equiv t_{i+1} - t_i)$, is small compared with the time constants of the system and when $\underline{A}(t)$, $\underline{B}(t)$ and $\underline{D}(t)$ are approximately constant over the sampling interval, are

$$\underline{\Phi}_i \approx \exp\{\underline{A}(t)\tau\} \approx \underline{I} + \underline{A}(t)\tau,$$

$$\underline{\Psi}_i \approx \underline{B}(t)\tau$$

$$\text{and } \underline{\Lambda}_i \approx \underline{D}(t)\tau,$$

where $\underline{I}$ is the unit matrix.

$$\text{Also } \underline{C}(t) = \underline{\Theta}_i.$$

3.2 Problem Statement

The estimation problem consists of finding the instantaneous estimates, $\underline{x}_{opt}(t)$, of the random system states, $\underline{x}(t)$, at time t , given the measurement data $\underline{z}(t)$ and the system equations, Equations (3.1) and (3.2), which together satisfy the following:

- (i) $\underline{x}_{opt}(t)$ is unbiased, i.e.

$$E\{\underline{x}_{opt}(t)\} = E\{\underline{x}(t)\}.$$

- and (ii) $\underline{x}_{opt}(t)$ is "optimal" in the sense that the expected value of the square of the error is minimised, i.e.

$$E\{[\underline{x}(t) - \underline{x}_{opt}(t)][\underline{x}(t) - \underline{x}_{opt}(t)]^T\} \text{ is a minimum.}$$

The estimation procedure is considered to comprise:

- (i) a prediction process where forecasts of the states of the system and their variances are obtained from previous best estimates

- and (ii) a filtering, or corrector, process that determines the best estimates of the forecast values and their variances by suitably combining these forecasts with measurements.

Some physical insight into the Kalman filter is possible when various extreme assumptions are made about the system and the levels of process noise and measurement noise present.

Consider the hypothetical case in which measurement noise is zero and in which the initial conditions of the system are known error free. In such a case the measurements provide error free estimates of the state(s) of a system and so it should be expected that these measurements alone are used by the filter to predict the system response. Consequently the Kalman gains should take values that eliminate the predicted state estimates from the Kalman algorithm.

Consider the equally hypothetical case when the system is absolutely free of process noise but the measurements are subject to instrumentation error. Assuming error free initial conditions it should now be expected that the predictions alone are used by the filter to predict the system response. The Kalman gains in this case take values that eliminate the measurements from the Kalman algorithm.

In both the cases cited above the error variance predictions of the system state(s) will be zero. Also, in the above cases, when the initial conditions have been chosen or evaluated erroneously, the estimated system states will, as the filter operates in time, align themselves with the true states to give steady-state error free estimates.

Other cases can be envisaged, for example a case when the process noise is large compared with the measurement noise. This will result in Kalman gains that heavily favour the measurements in preference to the predictions.

4. DERIVATION OF THE UNI-DIMENSIONAL KALMAN FILTER EQUATIONS

This section considers the derivation of both the discrete and continuous Kalman filter equations for a single state variable.

4.1 Discrete Kalman Filter

Simplification of Equations (3.3) and (3.4) for the case of a single state variable gives

$$x_{i+1} = \phi_i x_i + \lambda_i w_i \quad (4.1)$$

and
$$z_{i+1} = \theta_i x_{i+1} + v_{i+1}, \quad (4.2)$$

where for the sake of simplicity the deterministic input, u_i , is assumed zero. The process (or system) noise, w_i , and the measurement noise, v_i , are assumed to have zero means and variances of σ_w^2 and σ_v^2 , respectively. These variances, in general, may also be functions of time.

The initial state of the system, $x(t_0)$, is expressed by the following mean and variance values which are assumed known or otherwise available:

$$E\{x(t_0)\} = (x_{opt})_0$$

and
$$E\{[x(t_0) - (x_{opt})_0]^2\} = (\sigma_{opt})_0^2.$$

These known estimates of the state and its variance at time t_0 are used as starting values for the following procedure to obtain updated optimal estimates at time t_1 . Repeated application of the procedure provides the variables for all values of t_i at which measurements are available.

4.1.1 Prediction process

Suppose that an estimate of the state variable and its variance are both known at an initial time t_i , and are denoted by $(x_{opt})_0$ and $(\sigma_{opt})_0^2$, respectively.

From Equation (4.1) it is seen that the state changes in accordance with a difference equation that contains a white noise forcing function, w_i . A forecast of the state variable, denoted by the subscript f , can be derived assuming the absence of measurements as

$$(x_f)_{i+1} = \phi_i(x_{opt})_i. \quad (4.3)$$

The variance of this forecast value is obtained as follows.

Consider the error in $(x_{opt})_i$ to be ε_i ; then

$$(x_{opt})_i = x_i - \varepsilon_i. \quad (4.4)$$

The forecast value, $(x_f)_{i+1}$, of Equation (4.3) may be rewritten with the use of Equation (4.4) as

$$(x_f)_{i+1} = \phi_i x_i - \phi_i \varepsilon_i, \quad (4.5)$$

which, with the use of Equation (4.1) becomes

$$(x_f)_{i+1} = x_{i+1} - \lambda_i w_i - \phi_i \varepsilon_i. \quad (4.6)$$

Since x_{i+1} is the true value of the state variable at t_{i+1} , the last two terms on the right-hand side of Equation (4.6) provide a measure of the error of the forecast value, $(x_f)_{i+1}$.

From Equation (4.4) it is seen that the variance of ε_i is identical to the variance of $(x_{opt})_i$ and is given by $(\sigma_{opt})_i^2$. Thus the variance associated with the term $\{\phi_i \varepsilon_i\}$ is

$$V\{\phi_i \varepsilon_i\} = \phi_i^2 (\sigma_{opt})_i^2.$$

The variance of the term $\{\lambda_i w_i\}$ is

$$V\{\lambda_i w_i\} = \lambda_i^2 \sigma_w^2.$$

Consequently the variance of the forecast value of the state variable is

$$(\sigma_f)_{i+1}^2 = \phi_i^2 (\sigma_{opt})_i^2 + \lambda_i^2 \sigma_w^2, \quad (4.7)$$

where it has been assumed that two error terms are independent.