Initializing a Kalman Filter when Initial Values are Unknown

Hagander, Per

1973

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Hagander, P. (1973). *Initializing a Kalman Filter when Initial Values are Unknown*. (Technical Reports TFRT-7025). Department of Automatic Control, Lund Institute of Technology (LTH).

Total number of authors:
1

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

INITIALIZING A KALMAN FILTER WHEN INITIAL VALUES ARE UNKNOWN

Per Hagander

Report 7303 (B) February 1973
Lund Institute of Technology
Division of Automatic Control

INITIALIZING A KALMAN FILTER WHEN INITIAL VALUES ARE UNKNOWN.

P. Hagander

ABSTRACT.

The usual formulas for Kalman filters are not applicable when the initial values of the state are totally unknown. If the system is observable there still exists a unique estimate with a finite covariance from which the Kalman filter could be started.

For discrete time systems it may take some timesteps to achieve complete observability. During those steps there exists no unique estimate. The projection of the estimate on the unobservable subspace would lack a finite second moment, and could in fact be chosen arbitrarily, e.g. equal to zero. Formulas are deduced for these estimates and their covariance, and in the single output case it is shown how they could be obtained by two recursive equations. The consequence for linear stochastic control is also explained.

TABLE OF CONTENTS

1. INTRODUCTION.

State estimation for linear stochastic systems is a well established field, see e.g. [1]. The Kalman filter requires, however, a known statistic for the initial state. For an observable system it is possible to get an estimate of the state from the measurements of the output so that the Kalman filter could be started [2], even if the initial state is totally unknown. For discrete time systems the dimension of the observable subspace increases with time. A natural question is, what state estimates should be used before a system becomes completely observable.

Of course, one should use the available information for an estimate, although the projection on the unobservable subspace would lack a finite second moment. This component could in fact be chosen arbitrarily, for instance such that the norm of the estimate is minimized. A pseudoinverse could be used.

The consequences of unknown initial values for linear filtering are investigated in [2] both for continuous and discrete time. In [2] the duality with optimal control was used to get formulas for observable systems. Those results suggested a restructuring of the problem by splitting the estimate in two parts. It is thus shown in Section 2 how an estimate in the discrete time case can be obtained directly also when the system is not completely observable, i.e. during the first time steps.

Recursive equations for the start up are obtained in the single output case in Section 3 using results from recursive least squares identification [3, 4].

Some remarks on linear quadratic stochastic control and on cases with only a part of the initial state totally unknown concludes the report.

2. SPLITTING UP INTO TWO ESTIMATES.

Regard the system with totally unknown initial value x_0

$$\begin{cases} x(t+1) = \phi(t+1,t)x(t) + v(t), \quad x(0) = x_0 \\ y(t) = \theta(t)x(t) + e(t) \end{cases} \tag{1}$$

v and e are independent white noise with covariances R_1 and R_2, ($R_2 > 0$), and v and e are independent of x_0.

The solution to the start up of a Kalman filter with unknown initial values is given in [2]:

$$\hat{x}(t|t-1) = \hat{x}_\Pi(t|t-1) + \psi(t,0)M^{-1}(0,t)\lambda(-1,t) \tag{2}$$

where

$$\begin{cases} \hat{x}_\Pi(t+1|t) = \phi\hat{x}_\Pi(t|t-1) + K(t)[y(t) - \theta\hat{x}_\Pi(t|t-1)] \\ \hat{x}_\Pi(0|-1) = 0 \end{cases} \tag{3}$$

$$K(t) = \phi\Pi(t)\theta^T[\theta\Pi(t)\theta^T + R_2]^{-1} \tag{4}$$

$$\begin{cases} \Pi(t+1) = \phi\Pi(t)\phi^T + R_1 - K(t)\theta\Pi(t)\phi^T \\ \Pi(0) = 0 \end{cases} \tag{5}$$

$$\psi(t+1,t) = \phi(t+1,t) - K(t)\theta(t) \tag{6}$$

$$M(t_0,t_1) = \sum_{t=t_0}^{t_1-1} \psi^T(t,t_0)\theta^T(R_2 + \theta\Pi(t)\theta^T)^{-1}\theta\psi(t,t_0) \tag{7}$$

$$\lambda(t_0-1,t_1) = \sum_{t=t_0}^{t_1-1} \psi^T(t,t_0)\theta^T(R_2 + \theta\Pi(t)\theta^T)^{-1}[y(t) - \theta\hat{x}_\Pi(t|t-1)] \tag{8}$$

Note that (2) requires that the observability Gramian, M, is nonsingular, which is not the case during the first time steps.

The formula (2) suggests the splitting up of (1) into two parts:

$$\begin{cases} x_1(t+1) = \phi x_1(t) + v(t), \quad x_1(0) = 0 \\ y_1(t) = \theta x_1(t) + e(t) \end{cases} \tag{9}$$

$$\begin{cases} x_2(t+1) = \phi x_2(t), \quad x_2(0) = x_0 \\ y_2(t) = \theta x_2(t) \end{cases} \tag{10}$$

$$\begin{aligned} x &= x_1 + x_2 \\ y &= y_1 + y_2 \end{aligned} \tag{11}$$

Since K and Π used for $\hat{x}_{\Pi}$ are the same as in a Kalman filter for the system (9); i.e.

$$\begin{cases} \hat{x}_1(t+1|t) = \phi\hat{x}_1(t|t-1) + K(t)[y_1(t) - \theta\hat{x}_1(t|t-1)] \\ \hat{x}_1(0|-1) = 0 \end{cases} \tag{12}$$

It will be seen from the following that the second term of (2) is in fact the estimate of a quantity

$$z(t) = x_2(t) - [\hat{x}_{\Pi}(t|t-1) - \hat{x}_1(t|t-1)] \tag{13}$$

based on the measurements

$$\eta(t) = y(t) - \theta\hat{x}_{\Pi}(t|t-1) \tag{14}$$

In that way x_1 is estimated by $\hat{x}_{\Pi}$. The "x_2-estimate", $\hat{z}$, is corrected by the systematic error of the "x_1-estimate", $\hat{x}_{\Pi}$, so that the sum of the two estimates gives the minimum variance state estimate:

Theorem 2.1: The minimum variance estimate of the state $x(t)$ of (1) given Y_{t-1} is

$$\hat{x}(t|t-1) = \hat{x}_{\Pi}(t|t-1) + \hat{z}(t|t-1) \tag{15}$$

with $\hat{x}_{\Pi}$ defined by (3) and with $\hat{z}(t|t-1)$ being the minimum variance estimate of $z(t)$ given $\eta(s)$ up to $s = t-1$ (η_{t-1}). z and η are defined by (13) and (14).

Proof: The linearity gives

$$\hat{x}(t|t-1) = \hat{x}_1(t|Y_{t-1}) + \hat{x}_2(t|Y_{t-1})$$

where $\hat{x}_1(t|Y_{t-1})$ and $\hat{x}_2(t|Y_{t-1})$ are minimum variance estimates of $x_1(t)$ and $x_2(t)$ given Y_{t-1}.

It is easy to show how the vector

$$\eta_{t-1} = \begin{pmatrix} \eta(0) \\ \vdots \\ \eta(t-1) \end{pmatrix}$$

can be expressed as an invertible linear function of Y_{t-1}, so that the knowledge of η_{t-1} and Y_{t-1} is equivalent.

Eqn. (13) thus gives

$$\hat{z}(t|t-1) = \hat{x}_2(t|Y_{t-1}) - \hat{x}(t|t-1) + \widehat{[x_1(t|t-1)|Y_{t-1}]}$$

where the last term is the minimum variance estimate given Y_{t-1} of the minimum variance estimate of $x_1(t)$ given $Y_{1,t-1}$. Since $y_1(s)$ and $y_2(t)$ are independent, this term is equal to $\hat{x}_1(t|Y_{t-1})$, and (15) follows.□

In order to get nice formulas for $\hat{z}$, it can be proven that z and η are governed by a dynamic system:

Theorem 2.2: z and η defined by (13) and (14) satisfy the system

$$\begin{cases} z(t+1) = \psi(t+1,t)z(t), & z(0) = x_0 \\ \eta(t) = \theta z(t) + \varepsilon(t) \end{cases} \tag{17}$$

where ε is white noise with the covariance $(R_2+\theta\Pi\theta^T)$.

Proof: By direct use of the definitions of z, η and ψ:

$$\begin{aligned} z(t+1) &= x_2(t+1) - [\hat{x}_\Pi(t+1|t) - \hat{x}_1(t+1|t)] = \\ &= \phi[x_2(t) - \hat{x}_\Pi(t|t-1) + \hat{x}_1(t|t-1)] - \\ &\quad - K(t)[y(t) - y_1(t) - \theta\hat{x}_\Pi(t|t-1) + \theta\hat{x}_1(t|t-1)] = \\ &= \phi z(t) - K(t)[y_2(t) - \theta\hat{x}_\Pi(t|t-1) + \theta\hat{x}_1(t|t-1)] = \psi z(t) \end{aligned}$$

$$z(0) = x_2(0) - \hat{x}_\Pi(0|-1) + \hat{x}_1(0|-1) = x_0$$

$$\begin{aligned} \eta(t) &= y(t) - \theta\hat{x}_\Pi(t|t-1) = y_1(t) - \theta\hat{x}_1(t|t-1) + y_2(t) - \\ &\quad - \theta\hat{x}_\Pi(t|t-1) + \theta\hat{x}_1(t|t-1) = \tilde{y}_1(t|t-1) + \theta z(t) \end{aligned}$$

Define

$$\varepsilon(t) = \tilde{y}_1(t|t-1)$$

the Kalman filter error of (12), which is white noise with the covariance $R_2 + \theta\Pi\theta^T$.□

Having obtained this dynamic system (17) it is easy to find estimates $\hat{x}(t|t-1)$ also when the system is not completely observable, that is to generalize the formula (2).

Note that

$$\hat{z}(t|t-1) = \psi(t,0)\hat{z}(0|t-1)$$

where $\hat{z}(0|t-1)$ is the minimum variance estimate of the initial value $z(0) = x_0$. Using the matrix notation

$$A_t = \begin{bmatrix} \theta \\ \theta\psi(1,0) \\ \vdots \\ \theta\psi(t,0) \end{bmatrix} \qquad \eta_t = \begin{bmatrix} \eta(0) \\ \cdot \\ \cdot \\ \cdot \\ \eta(t) \end{bmatrix} \qquad \varepsilon_t = \begin{bmatrix} \varepsilon(0) \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon(t) \end{bmatrix} \tag{18}$$

the system (17) could be expressed as

$$\eta_t = A_t x_0 + \varepsilon_t \tag{19}$$

Let $R_{\varepsilon t}$ be the covariance matrix of ε_t. According to Theorem 2.2 $R_{\varepsilon t}$ is block-diagonal and invertible.

The well-known estimation theory, as for instance in [5], is now directly applicable to (19).

Corollary 2.1: Any minimum variance estimate $\hat{z}(0|t)$ of x_0, based on η_t must satisfy

$$A_t^T R_{\varepsilon_t}^{-1} A_t \hat{z}(0|t) = A_t^T R_{\varepsilon_t}^{-1} \eta_t \tag{20}$$

Its component in the nullspace of A_t could be chosen arbitrarily and has infinite error covariance. The orthogonal component, in the range space of A_t^T, is unique.□

Not that (20) always has at least one solution, since $A_t^T R_{\varepsilon_t}^{-1} \eta_t$ lies in the range space of $A_t^T R_{\varepsilon_t}^{-1} A_t$. The solution with minimal norm is obtained using the pseudoinverse.

Eqn. (20) could be reformulated as

$$\left[\sum_{s=0}^{t} \psi^T(s,0)\theta^T\left(\theta\Pi(s)\theta^T + R_2\right)^{-1}\theta\psi(s,0)\right]\hat{z}(0,t) =$$

$$= \sum_{s=0}^{t} \psi^T(s,0)\theta^T\left(\theta\Pi(s)\theta^T + R_2\right)^{-1}\eta(s)$$

or using the quantities M and λ defined by (7) and (8)

$$M(0,t)\hat{z}(0|t-1) = \lambda(-1,t) \tag{21}$$

The generalization of (2) could now be stated.

Corollary 2.2: The minimum variance state estimate $\hat{x}(t|t-1)$ for (1) is given by

$$\hat{x}(t|t-1) = \hat{x}_\Pi(t|t-1) + \psi(t,0)\hat{z}(0|t-1) =$$

$$= \hat{x}_\Pi(t|t-1) + \hat{z}(t|t-1)$$

with $\hat{x}_\Pi(t|t-1)$ from (3), $\psi(t,0)$ from (6) and $\hat{z}(0|t-1)$ fulfilling (21).□

Remark 1: Before the system becomes completely observable there is arbitrariness in $\hat{z}(0|t-1)$ for fulfilling (21). The natural choice is

Corollary 2.3: The minimum variance estimate $\hat{z}_m(0|t-1)$ with minimal norm is given by

$$\hat{z}_m(0|t-1) = (A_{t-1}^T R_{\varepsilon_{t-1}}^{-1} A_{t-1})^\dagger A_{t-1}^T R_{\varepsilon_{t-1}}^{-1} \eta_{t-1} =$$

$$= [M(0,t)]^\dagger \lambda(-1) = [D_{t-1} A_{t-1}]^\dagger D_{t-1} \eta_{t-1} \qquad (22)$$

with

$$D_{t-1} = R_{\varepsilon_{t-1}}^{-1/2}$$

and using a formula from [4] for the last expression.

Remark 2: All minimum variance estimates $\hat{z}(0|t-1)$ can be written

$$\hat{z}(0|t-1) = \hat{z}_m(0|t-1) + \gamma(0|t-1)$$

where $\gamma(0|t-1)$ is an arbitrary vector in the null space of $M(0,t)$. $\hat{z}_m(0|t-1)$ is also the minimum variance estimate of the orthogonal projection of x_0 on the range space of $M(0,t)$, of $M^\dagger M x_0$.

<u>Remark 3</u>: The covariance of $\tilde{z}_m(0|t-1) = M^{\dagger}Mx_0 - \hat{z}_m(0|t-1)$ is $M^{\dagger}(0,t)$.

<u>Remark 4</u>: For an observable system the covariance of the errors $\tilde{z}(0|t-1) = x_0 - \hat{z}(0|t-1)$ and $\tilde{z}(t|t-1) = z(t) - \hat{z}(t|t-1)$ are

$$\operatorname{cov} \tilde{z}(0|t-1) = M^{-1}(0,t)$$

$$\operatorname{cov} \tilde{z}(t|t-1) = \psi(t,0)M^{-1}(0,t)\psi^{T}(t,0) = \Sigma(t)$$

<u>Remark 5</u>: Define $\tilde{z}_m(t|t-1)$ and $\hat{z}_m(t|t-1)$ by

$$\tilde{z}_m(t|t-1) = \psi(t,0)M^{\dagger}(0,t)M(0,t)x_0 - \Psi(t,0)\hat{z}_m(0|t-1) =$$

$$= \psi(t,0)M^{\dagger}(0,t)M(0,t)x_0 - \hat{z}_m(t|t-1)$$

then

$$\operatorname{cov} \tilde{z}_m(t|t-1) = \psi(t,0)M^{\dagger}(0,t)\psi^{T}(t,0) = \Sigma(t)$$

by extending the definition of $\Sigma(t)$.

3. RECURSIVE EQUATIONS.

The pseudoinverse in (22) could be evaluated recursively, see e.g. [4]. The formulas are simplified for time invariant single output systems, where only one measurement is added at a time. The dimension of the unobservable subspace is then decreased by one for each time-step until the whole state space is observable after n steps. Only the single output case will be treated in detail. The elaborate expressions by Cline [6] could be used for the multioutput case, but they do not seem to have any computational advantages.

Theorem 3.1: A minimum variance estimate of x_0 from (17) can be obtained by the recursions:(single output case):

$$\begin{cases} \hat{z}(0|t) = \hat{z}(0|t-1) + K_i(t)[\eta(t) - \theta\psi(t,0)\hat{z}(0|t-1)] \quad (i = 1,2) \\ \hat{z}(0|-1) = 0 \end{cases}$$

1) $(0 \leq t < n)$

$$K_1(t) = P_1(t)\psi^T(t,0)\theta^T\left(\theta\psi(t,0)P_1(t)\psi^T(t,0)\theta^T\right)^{-1}$$

$$\begin{cases} P_1(t+1) = P_1(t) - K_1(t)\theta\psi(t,0)P_1(t) \\ P_1(0) = I \end{cases}$$

2) $(t \geq n)$

$$K_2(t) = P_2(t)\psi^T(t,0)\theta^T[\theta\psi(t,0)P_2(t)\psi^T(t,0)\theta^T+\theta\Pi(t)\theta^T+R_2]^{-1}$$

$$\begin{cases} P_2(t+1) = P_2(t) - K_2(t)\theta\psi(t,0)P_2(t) \\ P_2(n) = M^{-1}(0,n) \end{cases}$$

Proof: Corollary 2.3 gives

$$\hat{z}(0|t) = (D_t A_t)^{\dagger} D_t \eta_t$$

and as long as A_t has full row rank

$$(D_t A_t)^{\dagger} = A_t^{\dagger} D_t^{-1}$$

so that

$$\hat{z}(0|t) = A_t^{\dagger} \eta_t$$

Rewrite

$$A_t = \begin{bmatrix} A_{t-1} \\ \theta\psi(t,0) \end{bmatrix} \qquad \eta_t = \begin{bmatrix} \eta_{t-1} \\ \eta(t) \end{bmatrix}$$

and use the formula [4] for the pseudoinverse, when adding a row while increasing the rank $(0 < t < n)$:

$$A_t^{\dagger} = [Q \mid q]$$

$$Q = \Big[A_{t-1}^{\dagger} - (I - A_{t-1}^{\dagger} A_{t-1})\psi^T(t,0)\theta^T\theta\psi(t,0)A_{t-1}^{\dagger} /$$

$$/\theta\psi(t,0)\big(I - A_{t-1}^{\dagger} A_{t-1}\big)\psi^T(t,0)\theta^T\Big]$$

$$q = \Big[(I - A_{t-1}^{\dagger} A_{t-1})\psi^T(t,0)\theta^T /$$

$$/\theta\psi(t,0)[I - A_{t-1}^{\dagger} A_{t-1}]\psi^T(t,0)\theta^T\Big]$$

Define

$$P_1(t) = (I - A_{t-1}^{\dagger} A_{t-1})$$

and

$$K_1(t) = P_1(t)\psi^T(t,0)\theta^T/\theta\psi(t,0)P_1(t)\psi^T(t,0)\theta^T$$

then

$$\hat{z}(0|t) = \hat{z}(0|t-1) + K_1(t)[\eta(t) - \theta\psi(t,0)\hat{z}(0|t-1)]$$

and

$$P_1(t+1) = I - A^\dagger_{t-1}A_{t-1} + K_1(t)[\theta\psi(t,0)A^\dagger_{t-1}A_{t-1}-\theta\psi(t,0)] =$$

$$= P_1(t) - K_1(t)\theta\psi(t,0)P_1(t)$$

$$P_1(1) = I - \theta^\dagger\theta$$

$$\hat{z}(0|0) = \theta^\dagger\eta(0) \qquad\qquad K_1(1) = \theta^\dagger$$

which proves the first part of the theorem.

For the second part use D_t defined in Corollary 2.3 and

$$D_tA_t = \begin{bmatrix} D_{t-1}A_{t-1} \\ d_t\theta\psi(t,0) \end{bmatrix}$$

thus

$$d_t^2 = (\theta\Pi(t)\theta^T + R_2)^{-1}$$

Use the formula for adding a row without increasing the rank ($t \geq n$) and drop the indices $t-1$:

$$\hat{z}(0|t) = (D_tA_t)^{\dagger}D_t\eta_t =$$

$$= (DA)^{\dagger}D\eta + \frac{d_t^2(DA)^{\dagger}(DA)^{\dagger T}\psi^T(t,0)\theta^T}{1 + d_t^2\theta\psi(t,0)(DA)^{\dagger}(DA)^{\dagger T}\psi^T(t,0)\theta^T} \cdot$$

$$\cdot\ [\eta(t) - \theta\psi(t,0)(DA)^{\dagger}D\eta] =$$

$$= \hat{z}(0|t-1) + \frac{(A^TD^2A)^{-1}\psi^T(t,0)\theta^T}{d_t^{-2} + \theta\psi(t,0)(A^TD^2A)^{-1}\psi^T(t,0)\theta^T} \cdot$$

$$\cdot\ [\eta(t) - \theta\psi(t,0)\hat{z}(0|t-1)]$$

with

$$P_2(t) = (A^TD^2A)^{-1} = M(0,t)$$

the rest of the proof is an application of the well-known inversion lemma.□

Remark 1: Note that this minimum variance estimate is equal to the mean square estimate for $0 \leq t < n$.

Remark 2: Note that $P_2(t)$ is the error covariance for $\hat{z}(0|t-1)$, $t \geq n$, cf. Remark 4 of Section 2.

It is now easy to obtain formulas for $\hat{z}(t|t-1) = \psi(t,0)\hat{z}(0|t-1)$.

Corollary 3.1: A minimum variance estimate $\hat{z}(t|t-1)$ for (17) can be obtained by the recursions.

$$\begin{cases} \hat{z}(t+1|t) = \psi(t+1,t)\hat{z}(t|t-1) + K_i(t)[\eta(t) - \theta\hat{z}(t|t-1)] \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad (i = 3,4) \\ \hat{z}(0|-1) = 0 \end{cases} \tag{23}$$

1) $(0 \leq t < n)$

$$K_3(t) = \psi(t+1,t)P_3(t)\theta^T\left(\theta P_3(t)\theta^T\right)^{-1} \tag{24}$$

$$\begin{cases} P_3(t+1) = \psi(t+1,t)P_3(t)\psi^T(t+1,t) - \\ \qquad\qquad - K_3(t)\theta P_3(t)\psi^T(t+1,t) \\ P_3(0) = I \end{cases} \tag{25}$$

2) $(t \geq n)$

$$K_4(t) = \psi(t+1,t)P_4(t)\theta^T\left(\theta P_4(t)\theta^T + \theta\Pi(t)\theta^T + R_2\right)^{-1}$$

$$P_4(t+1) = \psi(t+1,t)P_4(t)\psi^T(t+1,t) - K_4(t)\theta P_4(t)\psi^T(t+1,t)$$

$$P_4(n) = \psi(n,0)M^{-1}(0,n)\psi^T(n,0) \tag{26}$$

Proof: Use

$$P_3(t) = \psi(t,0)P_1(t)\psi^T(t,0)$$

and

$$P_4(t) = \psi(t,0)P_2(t)\psi^T(t,0) \qquad \square$$

Remark 3: The term $\left(\theta P_3(t)\theta^T\right)$ is nonzero for $t < n$, but $P_3(n) = \underset{\sim}{0}$.

Remark 4: Note that

$$P_4(t) = \Sigma(t) = \text{cov}\ \tilde{z}(t|t-1) \qquad (t \geq n)$$

For $t \geq n$, $\hat{x}(t+1|t)$ should be computed by the usual Kalman filter where

$$P(t) = \Pi(t) + \Sigma(t)$$

but for $0 \leq t < n$ two recursions are needed, one for $\hat{x}_\Pi$ and one for $\hat{z}$.

4. REMARKS ON LINEAR QUADRATIC STOCHASTIC CONTROL.

The loss function

$$J = E\ell = E \sum_{0}^{N-1} \left(x^T(t)Q_1x(t) + u^T(t)Q_2u(t)\right) +$$

$$+ x^T(N)Q_0x(N) \tag{27}$$

should be minimized with respect to $u(0), \ldots, u(N)$ subject to the constraint

$$\begin{cases} x(t+1) = \phi x(t) + \Gamma u(t) + v(t) \qquad x(0) = x_0 \\ y(t) = \theta x(t) + e(t) \end{cases} \tag{28}$$

where v and e is independent zero mean Gaussian white noise with covariance R_1 and R_2. The mean value in (27) should be taken with respect to the introduced statistics (v and e). The initial state is an unknown constant as before. The choice of $u(t)$ is restricted to a linear map of Y_{t-1}.

Rewrite J using an identity in [1]:

$$J = E\left\{x_0^TS(0)x_0 + \sum_{0}^{N-1} v^T(t)S(t+1)v(t) + \right.$$

$$+ \sum_{0}^{N-1} \left(u(t) + L(t)x(t)\right)^T [Q_2 + \Gamma^TS(t+1)\Gamma] \cdot$$

$$\left. \cdot \left(u(t) + L(t)x(t)\right)\right\} =$$

$$= x_0^TS(0)x_0 + \sum_{0}^{N-1} \operatorname{tr} R_1(t)S(t+1) +$$

$$+ E \sum_{0}^{N-1} (u+Lx)^T \left(Q_2+\Gamma^TS(t+1)\Gamma\right)(u+Lx)$$

where

$$L(t) = \left(Q_2 + \Gamma^T S(t+1)\Gamma\right)^{-1} \Gamma^T S(t+1)\Phi \tag{29}$$

and

$$\begin{cases} S(t) = \phi^T S(t+1)\phi + Q_1 - \phi^T S(t+1)\Gamma L(t) \\ S(N) = Q_0 \end{cases} \tag{30}$$

Then the minimum of J say $V(x_0)$ could be written

$$V(x_0) = x_0^T S(0) x_0 + \sum_0^{N-1} \operatorname{tr} R_1(t) S(t+1) +$$

$$+ \min_{u(0),\ldots,u(m-1)} \left\{ E \sum_0^{m-1} (u+Lx)^T (Q_2 + \Gamma^T S \Gamma)(u+Lx) + \right.$$

$$\left. + \min_{u(m),\ldots,u(N-1)} E\left[\sum_m^{N-1} (u+Lx)^T (Q_2 + \Gamma^T S \Gamma)(u+Lx)\right]\right\}$$

where m-1 is the first time the whole state space is observable. But as in [1] with the control $u(t) = - L(t) \cdot \hat{x}(t|t-1)$, $m \leq t \leq N-1$,

$$\min_{u(m),\ldots,u(N-1)} E\left[\sum_m^{N-1} (u+Lx)^T (Q_2 + \Gamma^T S \Gamma)(u+Lx)\right] =$$

$$= \sum_m^{N-1} \operatorname{tr} P(t) L^T (Q_2 + \Gamma^T S \Gamma) L$$

where $\hat{x}(t|t-1)$ is the linear minimum variance estimate and $P(t)$ its covariance. During the first m steps, however, it is not possible to estimate the state, only its component in the observable subspace. Most of the statistics is thus eliminated.

$$V(x_0) = x_0^T S(0) x_0 + \sum_{0}^{N-1} \operatorname{tr} R_1(t) S(t+1) +$$

$$+ \sum_{m}^{N-1} \operatorname{tr} P(t) L^T(t) \left(Q_2 + \Gamma^T S(t+1)\Gamma\right) L(t) +$$

$$+ \min_{u(0),\ldots,u(m-1)} E \sum_{0}^{m-1} \left(u(t) + L(t)x(t)\right)^T \cdot$$

$$\cdot \left(Q_2 + \Gamma^T S(t+1)\Gamma\right)\left(u(t) + L(t)x(t)\right)$$

Only the last term remains. In order to eliminate the remaining statistics rewrite $x(t)$ using the results in Section 2.

$$x(t) = x_1(t) + x_2(t) = \hat{x}_{\Pi}(t|t-1) + \tilde{x}_1(t|t-1) +$$

$$+ \hat{z}_m(t|t-1) + \tilde{z}_m(t|t-1) + \psi(t,0)(I-M^+M)x_0$$

Introduce

$$F(t) = \psi(t,0)[I - M^+(0,t)M(0,t)] \tag{31}$$

Thus

$$\min_{u(0),\ldots,u(m-1)} \sum_{0}^{m-1} E\{||u + L\hat{x}_{\Pi} + L\tilde{x}_1 + L\hat{z}_m + L\tilde{z}_m +$$

$$+ LFx_0||^2_{(Q_2+\Gamma^T S\Gamma)}\} =$$

$$= \min_{u(0),\ldots,u(m-1)} E \sum_{0}^{m-1} ||u + L\hat{x}_{\Pi} + L\hat{z}_m + LFx_0||^2_{(Q_2+\Gamma^T S\Gamma)} +$$

$$+ \sum_{0}^{m-1} tr\left[\left(\Pi(t) + \Sigma(t)\right)\left(Q_2 + \Gamma^T S(t+1)\Gamma\right)\right] =$$

$$= \min_{\mu(0),\ldots,\mu(m-1)} \sum_{0}^{m-1} \left|\left|\mu(t) + L'(t)x_0\right|\right|^2_{(Q_2+\Gamma^T S\Gamma)} +$$

$$+ \sum_{0}^{m-1} tr\left\{\left(\Pi(t) + \Sigma(t)\right)\left(Q_2 + \Gamma^T S(t+1)\Gamma\right)\right\}$$

The first equality follows after some manipulations from $E[\tilde{x}_1(t|t-1)|Y_{t-1}] = 0$, $E[\tilde{z}_m(t|t-1)|Y_{t-1}] = 0$, $E\tilde{z}_m(t|t-1) = 0$, $E\tilde{z}_m(t|t-1)\tilde{x}_1^T(t|t-1) = 0$ and the Remark 5 of Section 2. The second equality is a reformulation using the definitions

$$\begin{cases} \mu(t) = u(t) + L(t)\left(\hat{x}_\Pi(t|t-1) + \hat{z}_m(t|t-1)\right) & (32) \\ \\ L'(t) = L(t)F(t) & (33) \end{cases}$$

It now remains the minimization of

$$\sum_{0}^{m-1} \left|\left|\mu(t) + L'(t)x_0\right|\right|^2_{(Q_2+\Gamma^T S(t+1)\Gamma)}$$

with respect to $\mu(t)$, $0 \leq t \leq m-1$.

Every $\mu(t)$ could be written

$$\mu(t) = L'(t)a(t) + b(t)$$

where $b(t)$ is perpendicular to the range space of $L'(t)$ in the scalar product induced by $\left(Q_2 + \Gamma^T S(t+1)\Gamma\right)$. Thus for each term

$$\left|\left|L'(t)a(t) + b(t) + L'(t)x_0\right|\right|^2_{(Q_2+\Gamma^T S(t+1)\Gamma)} =$$

$$= \left|\left|a(t) + x_0\right|\right|^2_{L'(t)^T(Q_2+\Gamma^T S(t+1)\Gamma)L'(t)} + \left|\left|b(t)\right|\right|^2_{(Q_2+\Gamma^T S\Gamma)}$$

and the choice $a(t) = 0$, $b(t) = 0$ or $\mu(t) = 0$ means a min max choice. Any other choice would make the sum larger for some x_0.

Summarize:

Theorem 4.1: The loss function J, (27), is minimized for the system (28) by

$$u(t) = \begin{cases} - L(t)[\hat{x}_\Pi(t|t-1) + \hat{z}_m(t|t-1)] & 0 \leq t < m \\ - L(t)\hat{x}(t|t-1) & t \geq m \end{cases}$$

with $L(t)$ from (29), $\hat{x}_\Pi(t|t-1)$ from (3) and $\hat{z}_m(t|t-1)$ from (22), giving the loss

$$V(x_0) = x_0^T S(0) x_0 + \sum_0^{m-1} x_0^T L'^T(t)\left(Q_2 + \Gamma^T S(t+1)\Gamma\right)L'(t)x_0 +$$

$$+ \sum_0^{N-1} \operatorname{tr} R_1(t)S(t+1) + \sum_0^{N-1} \operatorname{tr} P(t)\left(Q_2 + \Gamma^T S(t+1)\Gamma\right)$$

with $S(t)$ from (30), $L'(t)$ from (31) and (33) and $P(t) = \Pi(t) + \Sigma(t)$. Minimization is done for the worst possible x_0. □

5. ONLY SOME PART OF THE INITIAL VALUE TOTALLY UNKNOWN.

It is possible to prove analogous theorems for the generalisation that the initial state has a known statistic in a subspace, while it is totally unknown in the rest of the state space

$$x_0 = x_0^1 + x_0^2$$

where x_0^1 has mean value m and covariance R_0^1 and where x_0^2 is restricted to a subspace V but otherwise totally unknown. Let V be spanned by the row vectors of N.

Split up the system (1) by

$$\begin{cases} x_1(t+1) = \phi x_1(t) + v(t) & x_1(0) = x_0^1 \\ y_1(t) = \theta x_1(t) + e(t) & \end{cases} \qquad (9a)$$

$$\begin{cases} x_2(t+1) = \phi x_2(t) & x_2(0) = x_0^2 \\ y_2(t) = \theta x_2(t) & \end{cases} \qquad (10a)$$

Let the starting values of $\hat{x}_\Pi$ and Π be

$$\hat{x}_\Pi(0|-1) = m$$

$$\Pi(0) = R_0^1$$

and thus for z

$$z(0) = x_0^2 = N^T\xi$$

Then the minimum variance estimate of z(0), with minimal norm, is given by

$$\hat{z}(0|t-1) = N^T\left(NM(0,t)N^T\right)^{+}N\lambda(-1,t) \tag{22a}$$

and

$$\hat{x}(t|t-1) = \hat{x}_{\Pi}(t|t-1) + \psi(t,t-1)\hat{z}(0|t-1) \tag{16a}$$

When V is observable

$$\operatorname{cov} \tilde{z}(t|t-1) = \Sigma(t) = \psi(t,0)N^T\left(NM(0,t)N^T\right)^{-1}N\psi^T(t,0)$$

The recursive equations for this case suffer from the same complication as in the multioutput case. It must be decided, whether a certain measurement decreases the dimension of the intersection between V and the unobservable subspace, or not, in order to calculate the correct filter gain. It is therefore easier to solve (22a) in each time-step until the whole V becomes observable. Then the usual Kalman filter should be started with $P(t) = \Pi(t) + \Sigma(t)$.

6. CONCLUSIONS.

To start a Kalman filter the initial state is usually assumed to have a known statistic. This is, however, often unrealistic. It was here shown how the discrete time filter should be started, when the initial state, or part of it, is totally unknown.

Two filters are needed for $t < n$

$$\hat{x}(t+1|t) = \hat{x}_{\Pi}(t+1|t) + \hat{z}(t+1|t) \tag{15}$$

$$\begin{cases} \hat{x}_{\Pi}(t+1|t) = \phi\hat{x}_{\Pi}(t|t-1) + K(t)[y(t) - \theta\hat{x}_{\Pi}(t|t-1)] = \\ \qquad = \psi(t+1,t)\hat{x}_{\Pi}(t|t-1) + K(t)y(t) \\ \hat{x}_{\Pi}(0|-1) = 0 \end{cases} \tag{3}$$

$$\begin{cases} \hat{z}(t+1|t) = \psi(t+1|t)\hat{z}(t|t-1) + K_3(t)[y(t) - \theta\hat{x}_{\Pi}(t|t-1) - \\ \qquad - \theta\hat{z}(t|t-1)] \\ \hat{z}(0|-1) = 0 \end{cases} \tag{23}$$

where

$$\psi(t+1,t) = \phi(t+1,t) - K(t)\theta \tag{6}$$

and the gains K and K_3 are obtained from two different Riccati equations:

$$K(t) = \phi\Pi(t)\theta^T[\theta\Pi(t)\theta^T + R_2]^{-1} \tag{4}$$

$$K_3(t) = \psi P_3(t)\theta^T[\theta P_3(t)\theta^T]^{-1} \tag{24}$$

$$\begin{cases} \Pi(t+1) = \phi\Pi(t)\phi^T + R_1 - K(t)\theta\Pi(t)\phi^T \\ \Pi(0) = 0 \end{cases} \tag{5}$$

$$\begin{cases} P_3(t+1) = \psi P_3(t)\psi^T - K_3(t)\theta P_3(t)\psi^T \\ P_3(0) = I \end{cases} \tag{25}$$

while a usual Kalman filter could be started from $t = n$ with

$$\hat{x}(n|n-1) = \hat{x}_\Pi(n|n-1) + \hat{z}(n|n-1)$$

$$P(n) = \Pi(n) + \Sigma(n)$$

where

$$\Sigma(n) = \psi(n,0)\left[\sum_0^{n-1} \psi^T(s,0)\theta^T\left(R_2 + \theta\Pi(s)\theta^T\right)^{-1}\psi(s,0)\right]^{-1}\psi^T(n,0)$$

Note also that $P_3(n) = 0$.

In the multioutput case and in the case with the initial value restricted to a subspace of the state space, the recursions (23) and (25) are more complicated and it is better to solve $\hat{z}(t+1|t)$ directly from

$$\hat{z}(t+1|t) = \psi(t+1,t)\hat{z}(0|t)$$

and

7. REFERENCES.

[1] Åström, K.J.: Introduction to Stochastic Control Theory, Academic Press, New York, 1970.

[2] Hagander, P.: Linear Filtering with Unknown Initial Values. Report 7219 (B), Lund Institute of Technology, Division of Automatic Control (Aug. 1972).

[3] Åström, K.J.: Lectures on the Identification Problem - The Least Squares Method. Report 6806, Lund Institute of Technology, Division of Automatic Control (Sept. 1968).

[4] Söderström, T.: Notes on Pseudoinverses. Application to Identification. Report 7003, Lund Institute of Technology, Division of Automatic Control (July 1970).

[5] Luenberger, D.G.: Optimization by Vector Space Methods, John Wiley, New York, 1969.

[6] Cline, R.E.: Representation for the Generalized Inverse of Sums of Matrices. SIAM. J. Numer. Anal., Serie B, 2 (1965), 99 - 114.