An introduction to Kalman filter

Joel Le Roux, University of Nice leroux@essi.fr

3 janvier 2003

Table des matières

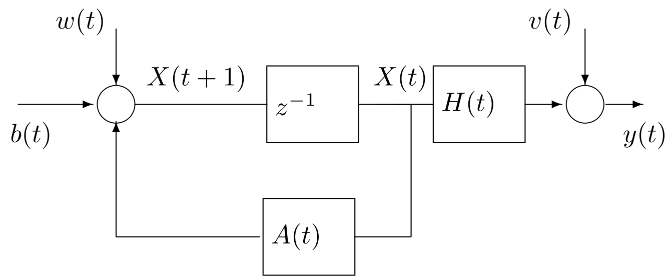

Figure 1: The process to by analysed by the Kalman filter

1 Introduction

In his 1960 famous publication (*"A new approach to linear filtering and prediction problems"*, Trans. ASME J. Basic Engineering., vol 82, March 1960, pp 34-45), Rudolf Kalman based the construction of the state estimation filter on probability theory, and more specifically, on the properties of conditional Gaussian random variables. The criterion he proposed to minimize is the state vector covariance norm, yielding to the classical recursion : the new state estimate is deduced from the previous estimation by addition of a correction term proportional to the prediction error (or the innovation of the measured signal).

If one tries to explain this remarkable and elegant construction to students having not a sufficient background in probability theory, there is an inherent difficulty (it is perhaps preferable to speak of a quite paradoxal aspect): its understanding requires a good knowledge of conditional gaussian random variables, while in the simple and efficient final formulation, the corresponding intermediate steps are not visible. In this presentation, I give a first construction based on probability theory in simple cases where Kalman approach is quite easy to follow : assuming the linearity of the predictor and avoiding the construction in the case of gaussian variables (this formulation based on linearity is mentionnend in Kalman's paper, but not developped, probably because considered as obvious by him). Then I propose a construction that is completely deterministic, (but as a consequence rather unnatural and inelegant ...) Perhaps some readers will be interested by this alternative construction?

In this deterministic construction, we consider two cases : in the first case, we start from a criterion based on the minimization of the prediction error ; we see in the final expressions that this formulation yields to the Kalman filter in the case where there is an observation noise but no control noise. In the second case, we modify the criterion by adding a penalty term in order to obtain a formula taking into account the control noise as well as the measurement noise ; however this second construction is quite artificial.

2 The Kalman filter

The state evolution is given by (cf. fig. 1)

$$X(t+1) = A(t)X(t) + b(t) + w(t), \tag{1}$$

where X is the state vector we intend to estimate, $A(t)$ is the known square transition matrix of the process. The control $b(t)$ is given and there is a zero mean process noise $w(t)$ with known covariance $r^w(t)$. The measured vector $y(t)$ is given by the measurement equation :

$$y(t) = H(t)X(t) + v(t). \tag{2}$$

$H(t)$ is the rectangular measurement matrix, $v(t)$ is the zero mean measurement noise, of known covariance $r^v(t)$. The dimension of $w(t)$ is the dimension of $x(t)$; the dimension of $v(t)$ is the dimension of $y(t)$. The covariance of the state vector $X(t)$ is

$$P(t) = E\left[\,(X(t) - E[X(t)])\left(X^T(t) - E[X^T(t)]\right)\right], \tag{3}$$

where X^T is the transpose (possibly conjugate) of X. The purpose of the Kalman filter is to deduce from $y(t)$ the vector $X(t)$ whose covariance matrix has the lowest norm (its trace). The steps of the estimation are the following:

• Prediction of the state $X(t)$:

$$X_{t+1/t} = A(t)X(t) + b(t); \tag{4}$$

• Intermediate update of the state covariance matrix that takes into account the evolution given by the process transition :

$$P_{t+1/t} = A(t)P(t)A^T(t) + r^w(t); \tag{5}$$

• Computation of the optimal gain :

$$K_{t+1} = P_{t+1/t}H^T(t+1)\big(H(t+1)P_{t+1/t}H^T(t+1) + r^v(t+1)\big)^{-1}; \tag{6}$$

this optimal gain depends on the statistical characteristics of the measurement noise, but it does not take the measures into account : it may be computed *a priori*.

• Update of the state covariance matrix :

$$\begin{aligned} P(t+1) = {} & P_{t+1/t} \\ & -P_{t+1/t}H^T(t+1)\big(H(t+1)P_{t+1/t}H^T(t+1) + r^v(t+1)\big)^{-1}H(t+1)P_{t+1/t}, \end{aligned} \tag{7}$$

or, expressed as a function of K_{t+1}

$$P(t+1) \quad = \quad [I - K_{t+1}H(t+1)]P_{t+1/t}; \tag{8}$$

• Computation of the new estimate of the state :

$$X(t+1) = X_{t+1/t} + K_{t+1}[y(t+1) - H(t+1)X_{t+1/t}]. \tag{9}$$

2.1 A simple case : estimation of a scalar

We consider the following problem : In order to estimate a constant m, we do several measurements on it ; let y be one of these measurements. y is a random variable with average m and fluctuations v :

$$y = m + v. \tag{10}$$

The zero mean noise v has variance σ_v^2. We will perform a recursive estimation of m : we suppose that we have a first biasless estimation of m in the form of a random variable x_0 :

$$x_0 = m + w. \tag{11}$$

The zero mean noise w has variance σ_w^2.

We intend to compute a new estimate x_1 with the following form :

$$x_1 = x_0 + k(y - x_0). \tag{12}$$

The specific chararacteristics of this correction are

• The new estimate is a linear function of the previous estimate and of the measurement ;

• The previous and the new estimators have no bias, or equivalently : if the measure y is perfectly predicted by the previous estimate x_0, this implies that it not necessary to correct this estimate as shows eq. (12) ; We note that the average of the new estimate is m : in replacing x_0 and y by their values :

$$x_1 = m + w + k(m + v - m - w), \tag{13}$$

$$x_1 = m + w + k(v - w), \tag{14}$$

$$E(x_1) = m. \tag{15}$$

We compute the variance of the new estimator assuming that the noise v is independant of x_0 ; so the variance of x_1 is

$$\sigma_1^2 = E(x_1 - m)^2 = E[w + k(v - w)]^2, \tag{16}$$

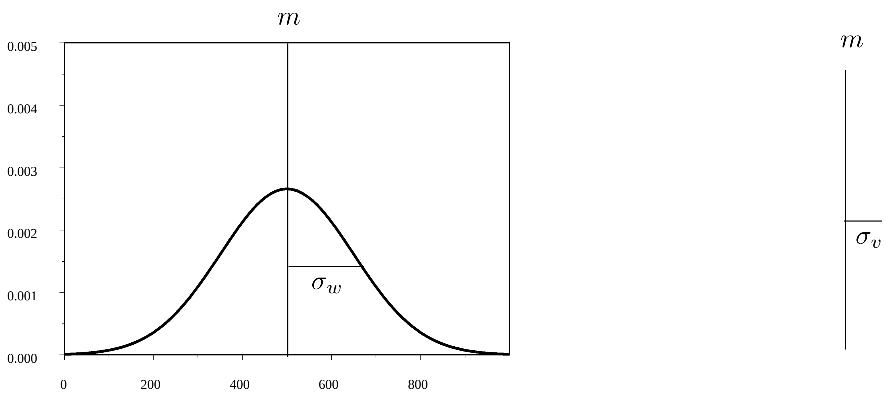

Figure 2: Two biasless estimators with variances σ_w^2 and σ_v^2; we look for a linear combination of these estimators giving a new estimator with lowest variance

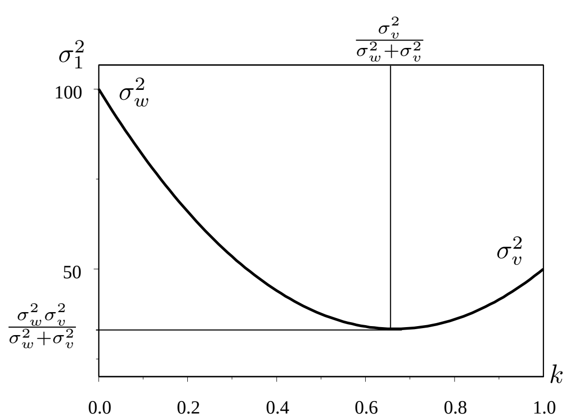

Figure 3: Variance of the new estimator as a function of k

$$\sigma_1^2 = (1-k)^2 E[w]^2 + k^2 E[v]^2, \tag{17}$$

$$\sigma_1^2 = (1-k)^2 \sigma_w^2 + k^2 \sigma_v^2. \tag{18}$$

It is reasonable to look for the estimate x_1 with lower variance and to compute the corresponding k. For this purpose we write

$$\sigma_1^2 = (1 - 2k + k^2)\sigma_w^2 + k^2\sigma_v^2, \tag{19}$$

$$\sigma_1^2 = \sigma_w^2 - 2k\sigma_w^2 + k^2(\sigma_w^2 + \sigma_v^2), \tag{20}$$

where we exhibit a constant term and a quadratic function of k

$$\sigma_1^2 = \sigma_w^2 - \beta^2 + \left(\beta - k\sqrt{\sigma_w^2 + \sigma_v^2}\right)^2, \tag{21}$$

$$\sigma_1^2 = \sigma_w^2 - 2\beta k\sqrt{\sigma_w^2 + \sigma_v^2} + k^2(\sigma_w^2 + \sigma_v^2). \tag{22}$$

Consequently,

$$\beta = \frac{\sigma_w^2}{\sqrt{\sigma_w^2 + \sigma_v^2}}, \tag{23}$$

$$\sigma_1^2 = \sigma_w^2 - \frac{\sigma_w^4}{\sigma_w^2 + \sigma_v^2} + \left(\frac{\sigma_w^2}{\sqrt{\sigma_w^2 + \sigma_v^2}} - k\sqrt{\sigma_w^2 + \sigma_v^2}\right)^2. \tag{24}$$

The minimum of σ_1^2 is obtained for

$$k = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_v^2}. \tag{25}$$

The value of this minimum is

$$\sigma_1^2 = \sigma_w^2 - \frac{\sigma_w^4}{\sigma_w^2 + \sigma_v^2}, \tag{26}$$

$$\sigma_1^2 = \frac{\sigma_w^2\sigma_v^2}{\sigma_w^2 + \sigma_v^2}, \tag{27}$$

$$\frac{1}{\sigma_1^2} = \frac{1}{\sigma_w^2} + \frac{1}{\sigma_v^2}. \tag{28}$$

We note that when the variance σ_w^2 of the previous estimate is very large, the new estimate reduces to the new measure, and its variance is σ_v^2 : so, when initializing the Kalman filter, it is reasonable to start from an uncertain initial state with large covariance if there is no *a priori* knowledge on the initial variance of the estimator of m. (fig. 4).

This simple case, which is quite easy to follow, gives the basic idea of the Kalman filter : the generalisation to the vectorial case and the introduction of the transition matrix A and of the control signal and noises are straightforwards.

2.2 State variance minimization in the case of vectorial data

In probabilistic terms, the problem considered by Kalman is the following : we have a biasless estimator of a vector X, we know its covariance ; we perform a new measure Y linearly dependant of X ; this measure has no bias and a given covariance ; how can we use this new measure in order to correct X by a linear term so that the new estimate is also biasless and has minimal variance?

The fact that the vector X can be interpreted as a state is important in applications in automatic control and filtering. However, this interpretation is not used in the computation of the optimal estimator. If we admit that the form of the Kalman filter is pertinent : that is the new vector state estimate is a biasless linear combination of the previous state estimate and of the the prediction error ; then the identification of the minimum covariance solution is similar to the formula given above in the scalar case.

The state estimate is x_0 (x_0 is a vector) :

$$x_0 = m + n, \tag{29}$$

m is the true value of the state (no bias) and n is the zero mean estimation noise .

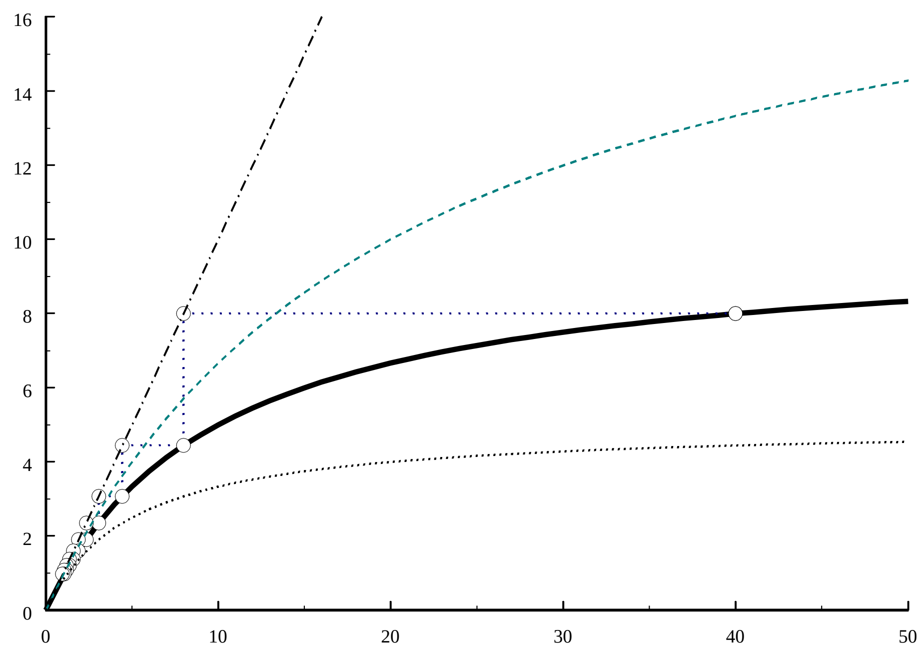

Figure 4: Evolution de σ_1^2 en fonction de σ_w^2 for different values of σ_v^2 in the scalar case.

The prediction of the measure y (y is also a vector, H is a rectangular matrix, since in general there are less measures than state components, then the dimension of y is less than the dimension of x) :

$$y = Hm + v \tag{30}$$

v is a noise independant of m and of n

We write the new state estimation in the form :

$$x_1 = x_0 + K(y_1 - Hx_0) \tag{31}$$

We look for a linear biasless estimator ; or equivalently : if we assume that the state estimate must not be corrected $x_1 = x_0$ if the prediction error $(y_1 - Hx_0)$ vanishes, then it is always possible to write x_1 in such a form : the mean of x_1 is equal to m as the mean of x_0 :

$$x_1 - m = x_0 - m + K(y_1 - Hm - H(x_0 - m)), \tag{32}$$

$$x_1 - m = x_0 - m + K(v - Hn), \tag{33}$$

$$E(x_1) = E(x_0) = m. \tag{34}$$

We assume that the measurement noise v is independant of the estimation error n (v does not appear in the computation of x_0).

Its covariance is

$$P_1 = E\left[(n - KHn + Kv)(n - KHn + Kkv)^T\right], \tag{35}$$

$$P_1 = (I - KH)P_0(I - KH)^T + KVK^T. \tag{36}$$

In order to minimize the norm (the trace) of the positive covariance P_1, it is useful to write it in the form

$$P_1 = P_0 - KHP_0 - P_0h^TK^T + K(HP_0H^T + V)K^T \tag{37}$$

and exhibit a sum of squares, so that only one of these squares will depend on the gain K we look for (this is one of the essential points of the original proof of R. Kalman, that was also present in the optimal filtering of N. Wiener.)

$$P_1 = P_0 - \alpha\alpha^T + \big(\alpha - K(HP_0H^T + V)^{1/2}\big)\big(\alpha - K(HP_0H^T + V)^{1/2}\big)^T \tag{38}$$

where we assume that we can factorize $HP_0H^T + V$ in

$$HP_0H^T + V = (HP_0H^T + V)^{1/2}(HP_0H^T + Q)^{T/2} \tag{39}$$

These factors may be triangular, but any other factorization could be used. The value of α is then

$$\alpha(HP_0H^T + V)^{T/2}K^T = P_0H^TK^T \tag{40}$$

$$\alpha = P_0H^T(HP_0H^T + V)^{-T/2} \tag{41}$$

We choose K yielding the minimum of P_1 : it is obtained when the third term in eq. (38) vanishes, since it is the only one depending of K and definite non negative. So,

$$K = \alpha(HP_0H^T + V)^{-1/2}, \tag{42}$$

$$K = P_0H^T(HP_0H^T + V)^{-T/2}(HP_0H^T + V)^{-1/2}, \tag{43}$$

$$K = P_0H^T(HP_0H^T + V)^{-1}. \tag{44}$$

and the covariance with minimal norm is

$$P_1 = P_0 - P_0H^T(HP_0H^T + V)^{-1}HP_0 \tag{45}$$

If instead of

$$x_1 = m + n, \tag{46}$$

the new state writes

$$x_1 = Am + An' + w, \tag{47}$$

where w is a control noise independant of the estimation error n', then in the previous equations P_0 is replaced by $(AP_0A^T + R)$ and the Kalman filter equations become

$$\boxed{K = (AP_0A^T + R)H^T[H(AP_0A^T + R)H^T + V]^{-1}} \tag{48}$$

$$\boxed{P_1 = (AP_0A^T + R) - (AP_0A^T + R)H^T[H(AP_0A^T + R)H^T + V]^{-1}H(AP_0A^T + R)} \tag{49}$$

In the case where the transition equation contains a nonlinear transformation of x_t instead of Ax_t, one has to replace the factor $(AP_0A^T + R)$ by a term taking into account the nonlinearity.

A remark on nonlinear extensions : S. J. Julier et J. K. Uhlmann (*A New Extension of the Kalman Filter to Nonlinear Systems. In Proc. of AeroSense: The 11th Int. Symp. on Aerospace/Defence Sensing, Simulation and Controls., 1997*) have proposed to analyse the effects of the nonlinearity on several points surrounding the estimated vector x in order to estimate the covariance of the nonlinearity output (fig. 5)

3 A deterministic construction of the Kalman filter

It is not always obvious (at least for me...) to see the link between the rather abstract probabilistic formulation of the previous section and prediction error minimization that is more familiar when dealing with actual engineering problems. Here I propose another approach to the construction of the Kalman filter where no probability theory is necessary.

In a first step, we suppose that there is only a measurement noise, and that the covariance of the control noise $w(t)$ is zero. The control noise will be added in a second step.

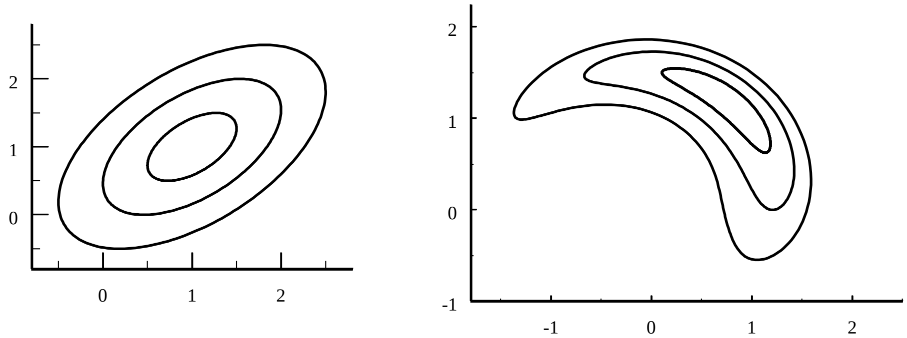

Figure 5: Transformation of a probability density by a nonlinearity ; in Julier's and Uhlman approach, the necessary information is a sufficient description of the probability density of the state so that its covariance can be estimated after the application of the nonlinearity .

3.1 The quadratic criterion to be minimized in the absence of process control noise and the corresponding solution

First, we note that in the case where the control noise is zero, there is no fundamental difference in estimating the state at time 0, at time t, at time T or at time $T+1$:

One changes from one of the estimates to another through a deterministic reversible formula : the state estimated at time $t+1$ can be deduced from the state that would have been estimated at time t by the transition equation :

$$x'(t+1) = A(t)x'(t) + b(t). \tag{50}$$

when t increases ; if t decreases, we have the corresponding computation by eq. (52) below.

At time t, the "prediction" error, that is the error between the measure $y(t)$ and its prediction from the state estimate $x'_T(t)$ is

$$\varepsilon(t) = y(t) - H'(t)x'_T(t), \tag{51}$$

with the recursive computation of $x'_T(t)$ given by

$$x'_T(t) = A^{-1}[x'_T(t+1) - b(t)], , \tag{52}$$

starting from

$$x'_T(T) = x(T).. \tag{53}$$

We look for x'_T that minimizes

$$\sum_{t=0}^{T} \varepsilon(t)^T C(t)\varepsilon(t) + \varepsilon(T+1)^T C(T+1)\varepsilon(T+1). \tag{54}$$

where

$$\varepsilon(T+1) = y(T+1) - H(T+1)\left[A(T)x'_T + b(T)\right] \tag{55}$$

$C(t)$ is a definite positive weighting matrix that will not appear explicitely in the sequel of the computations (for $t \leq T$) ; $C(T+1)$ is the weight of the new measurement error that will appear explicitely.

In a first step we intend to update x_T, which yields x'_{T+1} from which we shall deduce x_{T+1} by (50).

We look for an iterative solution in assuming that we know the solution x_T that minimizes

$$\sum_{t=0}^{T} \varepsilon(t)^T C(t)\varepsilon(t), \tag{56}$$

that is

$$\left[\sum_{t=0}^{T} H'_T(t)^T C(t) H'_T(t)\right] x(T) = \sum_{t=0}^{T} H'_T(t)^T C(t) y'(t). \tag{57}$$

In eq. (57), $H'_T(t)$ is given by

$$H'_T(t) = H(t)A^{-1}(t) \times \ldots \times A^{-1}(T). \tag{58}$$

In eq. (57) the $y'_T(t)$ would given by

$$y'_T(t) = y(t) - H(t)b'_T(t) \tag{59}$$

The $b'_T(t)$ are computed recursively

$$b'_T(t) = b(t) - A^{-1}(t)b'_T(t+1), \tag{60}$$

$$b'_T(T) = b(T). \tag{61}$$

Since we look for a recursive solution, these terms will not need to be computed explicitely.
We shall denote the matrices of eq. (57) in the form

$$P_T^{-1} x_T = Q_T. \tag{62}$$

As we assume that we know this solution, we do not need to compute explicitely the terms $\varepsilon(t)$ or $H'(t)$. The solution that minimizes (54) can be written

$$\begin{gathered}\left[P_T^{-1} + A^T(T)H^T(T+1)C(T+1)H(T+1)A(T)\right] x'_{T+1} = \\ Q_T + A^T(T)H^T(T+1)C(T+1)\left[y(T+1) - H(T+1)b(T)\right],\end{gathered} \tag{63}$$

that we shall rewrite in using the same notations as in eq. (62):

$$P_{T+1}^{-1} x'_{T+1} = Q_{T+1}, \tag{64}$$

with

$$P_{T+1}^{-1} = P_T^{-1} + A^T(T)H^T(T+1)C(T+1)H(T+1)A(T). \tag{65}$$

In the sequel, in order to lighten the expressions, we shall rewrite (63)

$$\left[P^{-1} + A^T H^T C H A\right] x'_{T+1} = Q + A^T H^T C\left[y - Hb\right] \tag{66}$$

3.2 Iterative expression of the solution

3.2.1 Recursion on the matrix P

According to the matrix inversion lemma applied to eq. (65), P_{T+1} can be written

$$P_{T+1} = \left[P^{-1} + A^T H^T C H A\right]^{-1} = P - PA^T H^T\left[C^{-1} + HAPA^T H^T\right]^{-1} HAP, \tag{67}$$

that we rewrite

$$P_{T+1} = P - PA^T H^T G H A P, \tag{68}$$

where we name

$$G = \left(C^{-1} + HAPA^T H^T\right)^{-1}. \tag{69}$$

One can left multiply (67) by A and right multiply by A^T, which yields a result that we shall use subsequently in the Kalman filter recursion :

$$\boxed{AP_{T+1}A^T = APA^T - APA^T H^T G H A P A^T.} \tag{70}$$

3.2.2 Modification of the expression of the solution

According to (63), the solution we look for is

$$x'_{T+1} = \left[P - PA^T H^T GHAP\right] \left[Q + A^T H^T C \left[y - Hb\right]\right] \tag{71}$$

or, in developping :

$$\begin{aligned} x'_{T+1} = & \left[PQ - PA^T H^T GHAPQ\right] \\ & + \Big(\left[P - PA^T H^T GHAP\right] A^T H^T C \left[y - Hb\right] \Big). \end{aligned} \tag{72}$$

In making x_T appear, according to (62)

$$\begin{aligned} x'_{T+1} = & \left[x_T - PA^T H^T GHAx_T\right] \\ & + \Big(\left[P - PA^T H^T GHAP\right] A^T H^T C \left[y - Hb\right] \Big). \end{aligned} \tag{73}$$

We develop the factor of $[y - Hb]$ that we name S

$$S = \left[P - PA^T H^T GHAP\right] A^T H^T C \tag{74}$$

in the second term of the sum (73). We can simplify the expression of S :

$$S = PA^T H^T C - PA^T H^T GHAPA^T H^T C. \tag{75}$$

We introduce artificially $0 = C(T+1)^{-1} - C(T+1)^{-1}$:

$$S = PA^T H^T C - \Big(PA^T H^T G \left(HAPA^T H^T + C^{-1} - C^{-1}\right) C \Big) \tag{76}$$

and we recognize G^{-1}

$$S = PA^T H^T C - PA^T H^T G \left(G^{-1} - C^{-1}\right) C \tag{77}$$

$$S = PA^T H^T C - PA^T H^T \left(I - GC^{-1}\right) C \tag{78}$$

$$S = PA^T H^T C - \left(PA^T H^T C - PA^T H^T GC^{-1} C\right) \tag{79}$$

$$S = PA^T H^T G \tag{80}$$

3.3 Final expression of the solution

So, the solution of (63) writes

$$x'_{T+1} = x_T - PA^T H^T GHAx_T + PA^T H^T G \left[y - Hb\right]. \tag{81}$$

$$x'_{T+1} = x_T - PA^T H^T G\big(HAx_T - y + Hb\big) \tag{82}$$

$$x'_{T+1} = x_T + PA^T H^T G\big(y - H(Ax_T + b)\big). \tag{83}$$

In the following iterations, the state to be memorized is no longer x'_{T+1} but instead

$$x_{T+1} = A(T) x'_{T+1} + b(t) \tag{84}$$

In reintroducing the notations taking account of time, and in replacing G by its expression (69)

$$G = \left(C^{-1} + HAPA^T H^T\right)^{-1}. \tag{85}$$

x_{T+1} writes

$$\boxed{x_{T+1} = Ax_T + b + APA^T H^T \left(C^{-1} + HAPA^T H^T\right)^{-1} \big(y - H(Ax_T + b)\big).} \tag{86}$$

3.4 Summary of the computations

So, we see that we obtain the steps of the recursion of the Kalman filter in the absence of control noise (rewriting of equations (70) and (86) that we write in using the classical decomposition with notations taking time into account :

• Prediction of the state transition before the correction due to the new measures

$$x_{T+1/T} = A(T)x_T + b(T); \tag{87}$$

• Intermediate update of the covariance matrix (remember that we have supposed that the control noise is zero)

$$P_{T+1/T} = A(T)P_T A^T(T); \tag{88}$$

In the case where we suppose the presence of a control noise of covariance r^w, we take this noise into account by modifying this formula : The importance of the correction is modified in considering that the prediction error is partly due to this control noise.

• Computation of the Kalman gain

$$K_{T+1} = P_{T+1/T}H^T(T+1)\big(C^{-1} + H(T+1)P_{T+1/T}H^T(T+1)\big)^{-1}; \tag{89}$$

• Update of the state covariance (equation (70))

$$\begin{aligned} P(T+1) &= P_{T+1/T} \\ &-P_{T+1/T}H^T(T+1)\Big(C(T+1)^{-1} + H(T+1)P_{T+1/T}H^T(T+1)\Big)^{-1}H(T+1)P^T_{T+1/T}; \end{aligned} \tag{90}$$

or

$$P(T+1) = \Big(I - K_{T+1}H(T+1)\Big)P^T_{T+1/T}; \tag{91}$$

• Update of the state estimate

$$x(T+1) = x_{T+1/T} + K_{T+1}\left[y(T+1) - H(T+1)x_{T+1/T}\right]. \tag{92}$$

3.5 Introduction of the penalty term in the criterion

If a control noise with covariance r^w is taken into account, the correction factor is modified : we assume that the error is partly due to this control noise and the optimal solution would be

$$\begin{aligned} x_{T+1} &= A_T x_T + b_T \\ &+(APA^T + R)H^T\left(C^{-1} + H(APA^T + R)H^T\right)^{-1}(y_{T+1} - (Ax_T + b)), \end{aligned} \tag{93}$$

The new expression of the matrix P at step $T+1$ being

$$\begin{aligned} P_{T+1} &= A_T P_T + R_T \\ &-(APA^T + R)H^T\left(C^{-1} + H(APA^T + R)H^T\right)^{-1}H(APA^T + R). \end{aligned} \tag{94}$$

We propose a modification of the criterion in order to obtain a solution of this form/

If we want to obtain a formula similar to that of the Kalman filter recursion where there is a control noise, we must replace the term APA^T by a term of the form $APA^T + R$, or P^{-1} by a term of the form $P^{-1} + Z$ in the equation (63)

$$\left(P^{-1} + Z + A^T H^T CHA\right)x'_{T+1} = Q_T + A^T H^T C(y - Hb). \tag{95}$$

to be solved. Z is symmetric, and $\left(P^{-1} + Z + A^T H^T CHA\right)$ must be definite positive

This introduction requires a modification of the criterion (54) that will be changed in

$$\sum_{t=0}^{T} \varepsilon(t)^T C(t)\varepsilon(t) + \varepsilon(T+1)^T C(T+1)\varepsilon(T+1) + (x'_{T+1})^T Z(T+1) x'_{T+1}. \tag{96}$$

The introduction of the penalty term $(x'_{T+1})^T Z(T+1) x'_{T+1}$ yields a solution of the form (95) and a construction where $A^T P A$ is replaced by $A^T P A + R$. However we have to establish the relationship between the expressions of $Z(T)$ and $R(T+1)$.

It is possible to write $P^{-1} + Z(T)$ in the form

$$P^{-1} + Z(T+1) = (A^{-T} R A^{-1} + P)^{-1}, \tag{97}$$

$$Z(T+1) = (A^{-T} R A^{-1} + P)^{-1} - P^{-1} \tag{98}$$

$$Z(T+1)(A^{-T} R A^{-1} + P) = I - P^{-1}(A^{-T} R A^{-1} + P) \tag{99}$$

$$Z(T+1)(A^{-T} R A^{-1} + P) = -P^{-1} A^{-T} R A^{-1} \tag{100}$$

$$Z(T+1) = -P^{-1} A^{-T} R A^{-1} (A^{-T} R A^{-1} + P)^{-1} \tag{101}$$

$$Z(T+1) = -P^{-1} A^{-T} R (R + A^T P A)^{-1} A^T \tag{102}$$

The scalar case example may be useful to have an idea of the relationship between Z and R,

$$Z = -\frac{R}{P(R + A^2 P)}, \tag{103}$$

We note that when R becomes very small or vanishes, Z is also very small or vanishes ; when R is very large, $P^{-1} + Z$ becomes very small.

The choice of the weighting function C and of the penalty term Z must be coherent with the formulation of the Kalman filter. C is interpretated as the inverse of the measurement noise covariance ; Z as the inverse of the control noise covariance.

This rather artificial construction