Initial Conditions for Kalman Filtering: Prior Knowledge Specification

EVGENIA SUZDALEVA
Department of Adaptive Systems
ÚTIA AV ČR
Pod vodárenskou věží 4, 18208 Prague
CZECH REPUBLIC

Abstract: The paper deals with a selection of the initial state for Kalman filtering. The prior knowledge about it can be highly uncertain. In practice the initial state mean and covariance are often chosen arbitrarily. The present paper considers the problem from the position of knowledge elicitation and proposes a methodology to extract the prior knowledge from available information by the respective processing in order to choose the adequate initial conditions. The suggested methodology is based on utilization of the conjugate prior distribution for models, belonging to the exponential family.

Key–Words: Kalman filtering, prior knowledge, state-space model, exponential family

1 Introduction

The paper describes the methodology of extraction of the prior knowledge about the initial state distribution, used for Kalman filtering, from available expert information. Kalman filter [1] can be well implemented due to good specification of the initial state and the assumed knowledge of variances of involved noises. The application domain of the work is the transportation control, where there is a lot of expert knowledge available, for instance, precision of sensors, limited range of involved signals, data from simulated or similar traffic systems or etc. It is necessary to translate this knowledge into the corresponding probability density functions (pdf) and use them to choose the adequate initial conditions for Kalman filtering. This task is addressed in the paper.

The research in this area is rather concerned with the input-output models. The methodology, proposed in [2], used a complicated procedure of weighting of the prior knowledge pieces. Subsequent research in this field [3] proved, however, that the underlying concept of fictitious data, used in [2], can be replaced by the more general, but less elaborated problem formulation as optimization under informational restriction. The paper [3] proposes a promising methodology of translation of the specific expert knowledge into the prior pdf, used by Bayesian estimation. The present paper extends and modifies this methodology for the state-space models.

The outline of the paper consists of the following parts. Basic facts about the models used and Bayesian state estimation are provided in Section 2. The main emphasis of the paper is on the technique of prior knowledge extraction, which is described in Section 3. The proposed technique is based on the use of the conjugate prior distribution for models, belonging to the exponential family. Due to this property of the models, belonging to the exponential family, one can obtain the *transformed pdf*, expressing the initial state, conditional on the available pair $(input, output)$ of the expert knowledge. The result of application of the technique gives the recursively calculated statistics for the specified form of the model.

The model is taken as the recursive function and proposed so that it would give the possibility to use the relation between the initial state only and a various number of available system inputs and outputs, provided by the experts.

Section 4 demonstrates the application of the proposed methodology for normal state-space model. Concluding remarks in Section 5 close the paper.

2 Preliminaries

The probabilistic description of the system, which state is to be estimated, is provided by the models in the form of the following pdfs.

2.1 Models

The *model of observation*

$$f\left(y_t | u_t, x_t\right), \quad (1)$$

relates the outputs y_t to the inputs u_t and state x_t at discrete time moments $t \in t^* \equiv \{1, \ldots, \mathring{t}\}$.
The *model of state evolution*

$$f\left(x_t | u_t, x_{t-1}\right), \quad (2)$$

describes time evolution of the state x_t.
The *model of control strategy*

$$f\left(u_t \middle| d^{\,t-1}\right), \quad (3)$$

describes, generally randomized, generating of inputs u_t based on $d^{\,t-1}$, where $d^{\,t} \equiv (y_t, u_t)$.

Proposition 1 (Assumptions) *It is assumed that neither the output y_t nor the state x_t depend on the past data $d^{\,t-1}$, and control strategies ignore the unobservable system states.*

The finite-dimensional system state has to be estimated and the system output has to be predicted. These operations call for application of Bayesian prediction and filtering.

2.2 Prediction and Filtering

Bayesian predictor of the output is given by the formula

$$f\left(y_t \middle| u_t, d^{\,t-1}\right) = \int f\left(y_t | u_t, x_t\right) f\left(x_t \middle| u_t, d^{\,t-1}\right) dx_t. \quad (4)$$

Bayesian filtering, estimating the state x_t, includes the following coupled formulas.
Data updating

$$f\left(x_t \middle| d^{\,t}\right) = \frac{f\left(y_t | u_t, x_t\right) f\left(x_t \middle| u_t, d^{\,t-1}\right)}{f\left(y_t \middle| u_t, d^{\,t-1}\right)}, \propto f\left(y_t | u_t, x_t\right) f\left(x_t \middle| u_t, d^{\,t-1}\right), \quad (5)$$

($\propto$ means proportionality) that incorporates the experience contained in the data $d^{\,t}$.
Time updating

$$f\left(x_{t+1} \middle| u_{t+1}, d^{\,t}\right) = \int f\left(x_{t+1} | u_{t+1}, x_t\right) f\left(x_t \middle| d^{\,t}\right) dx_t, \quad (6)$$

which fulfills the state prediction.

The filtering does not depend on the control strategy $\{f(u_t|d^{\,t-1})\}_{t\in t^*}$ but on the generated inputs only.

The application to Gaussian state-space model with Gaussian prior on x_0 and Gaussian observations provides Kalman filter. The prior pdf $f(x_0)$, that expresses the subjective prior knowledge on the initial state x_0, starts the recursions. The choice of the mentioned pdf is the main question of the paper.

3 Prior Knowledge Extraction

For application in traffic control, the paper deals with, the prior pdf $f(x_0)$ reflects the uncertain knowledge about initial length of the queue on the intersection. The expert knowledge about the intersection can include the corresponding system parameters, the specific traffic characteristics such as a saturated flow of an intersection lane, a turn rate, time of the green light (the system input), measurements of the sensors, etc. At the present paper it is assumed, that the experts provided the data, including the inputs and outputs of the traffic system.

Let the expert knowledge be described by the pdfs $f_{\tau^*} \equiv \{f(d^{\,\tau})\}_{\tau\in\tau^*}$, where τ^* is a finite set of time moments. The subscript τ relates to the quantities, expressing the prior knowledge, while t – to the observed ones. The distribution of the initial state x_0 should be chosen, taking into account this expert knowledge.

The methodology, proposed in [3], solves a similar problem for the system identification case with input-output models. Modifying it for Bayesian *state* estimation, the flat prior pdf $f(x_0)$ is suggested to be transformed into

$$f(x_0|f_{\tau^*}) = \frac{f(x_0)\exp[\Omega_{\tau^*}(x_0)]}{\int f(x_0)\exp[\Omega_{\tau^*}(x_0)]\,dx_0}, \quad (7)$$

where the form of the function $\Omega_{\tau^*}(x_0)$ depends on the cardinality of the set τ^*, denoted $\mathring{\tau}$. For example, for $\tau = 0$, i.e. for the pair of knowledge pieces with the zero index $f(y_0, u_0)$, provided by experts, the pdf (7) has a relatively simple form

$$f(x_0|f(y_0,u_0)) \propto f(x_0) \times \underbrace{\exp[\int f(y_0,u_0)\ln[\underbrace{f\left(y_0|\,u_0,x_0\right)}_{\mathcal{Z}_{\tau=0}(d^{\,\tau}|x_0)}]\,dy_0du_0]}_{\Omega_{\tau^*}(x_0)}. \quad (8)$$

The function $\mathcal{Z}_{\tau=0}(d^{\,\tau}|x_0) \equiv \mathcal{Z}(d^{\,0}|x_0)$ in (8) reflects the state-space model. As the only zero index pair of knowledge pieces is available, only the model of observation (1) is used here. The situation becomes more difficult, when the expert knowledge includes $f(y_0, y_1, u_0, u_1)$. The state-space model requires the following modified form of the function $\mathcal{Z}(d^{\,1}|x_0)$.

$$\begin{aligned} & f(y_0, y_1, u_0, u_1|x_0) \\ = & \; f(y_1, u_1|y_0, u_0, x_0)\, f(y_0, u_0|x_0), \\ = & \int f(y_1, x_1, u_1|y_0, u_0, x_0)\; dx_1 f(y_0, u_0|x_0), \\ = & \int f(y_1|x_1, u_1)\, f(x_1|u_1, x_0)\; dx_1 \\ \times & \; f(y_0, u_0|x_0), \\ = & \int f(y_1|x_1, u_1)\, f(x_1|u_1, x_0)\; dx_1 \\ \times & \; \underbrace{f(y_0|u_0, x_0)}_{\mathcal{Z}(d^{\,0}|x_0)}, \end{aligned} \tag{9}$$

which can be obtained with the help of operation of marginalization and Assumptions (1). Note, that all the pdfs here are known from the model (1)-(2). After integrating and some algebraic rearrangements such a form of the function $\mathcal{Z}(d^{\,1}|x_0)$ provides the initial state, conditional on the expert knowledge $f(y_0, u_0, y_1, u_1)$ only, while the state x_1 is being integrated out. It is clear, that the function $\Omega_{\tau^*}(x_0)$ is integrated over y_0, y_1, u_0 and u_1 in this case.

Recalculating the relation (9) for a subsequent τ, one can note, that the function $\mathcal{Z}(d^{\,\tau}|x_0)$ can be expressed recursively in the following way.

$$\begin{aligned} \mathcal{Z}(d^{\,\tau}|x_0) \;=\; & \int f(y_\tau|x_\tau, u_\tau) \\ \times & \; f(x_\tau|u_\tau, x_{\tau-1}) \\ \times & \; f(x_{\tau-1}|u_{\tau-1}, x_{\tau-2}) \ldots \\ \times & \; f(x_{\tau-i}|u_{\tau-i}, x_0)\; dx_\tau \ldots\, dx_{\tau-i} \\ \times & \; \mathcal{Z}(d^{\,\tau-1}|x_0), \end{aligned} \tag{10}$$

where $i = \tau - 1$, $\tau = 1, \ldots, \mathring{\tau}$. Therefore, pdf (7) can be rewritten as

$$\begin{aligned} f(x_0|d^{\,\tau}) \;\propto\; & f(x_0) \\ \times & \; \exp \underbrace{\left[\int f(d^{\,\tau}) \ln[\mathcal{Z}(d^{\,\tau}|x_0)]\, dd^{\,\tau} \right]}_{\Omega_{\tau^*}(x_0)}. \end{aligned} \tag{11}$$

The function $\Omega_{\tau^*}(x_0)$ has a simple form in the case, when the model belongs to the exponential family (Gaussian distribution does).

$$\mathcal{Z}_\tau(d^{\,\tau}|x_0) = A(x_0) \exp \langle B(d^{\,\tau}), C(x_0) \rangle, \tag{12}$$

where $A(x_0)$ is a non-negative scalar function, defined on x_0; $B(d^{\,\tau})$ and $C(x_0)$ are multivariate functions of compatible and finite dimensions; the functional $\langle \cdot, \cdot \rangle$ is linear in the first argument. In this case pdf (7) will get the following form

$$f(x_0|f_{\tau^*}) \;\propto\; f(x_0) A(x_0) \exp \langle \mathring{\tau} V, C(x_0) \rangle, \tag{13}$$

where the statistic V is

$$V \;\equiv\; \frac{1}{\mathring{\tau}} \sum_{t \in t^*} \int f(d^{\,\tau}) B(d^{\,\tau})\, dd^{\,\tau}. \tag{14}$$

Since both the model and the prior (flat) pdf belong to the exponential family, the conjugate prior distribution can be used. The utilization of the conjugate prior leads to keeping the same form for the transformed pdf $f(x_0|f_{\tau^*})$, because it will also belong to the exponential family in such a case. The conjugate prior pdf can be taken as

$$f(x_0) \;=\; \frac{A(x_0) \exp \left\langle \bar{V}, C(x_0) \right\rangle}{\int A(x_0) \exp \left\langle \bar{V}, C(x_0) \right\rangle dx_0}. \tag{15}$$

With such a conjugate prior pdf the transformed pdf $f(x_0|\tau^*)$ keeps this form with recursively calculated statistic

$$V_\tau \;=\; V_{\tau-1} + B(d^{\,\tau}),\; V_0 \equiv \bar{V} + \mathring{\tau} V. \tag{16}$$

The statistic V in (16) defines the pdf $f(x_0|\tau^*)$ in general case of Kalman filter, when the variances are supposed to be known. For estimation with the unknown variances the function $A(x_0)$ in the conjugate prior (15) becomes $A^{\bar{\nu}}(x_0)$, where $\bar{\nu}$ is a degree of freedom. In this case one more statistic should be defined for the transformed pdf $f(x_0|\tau^*)$, and it is recursively calculated as

$$\nu_\tau \;=\; \nu_{\tau-1} + 1,\; \nu_0 \equiv \bar{\nu} + \mathring{\tau}. \tag{17}$$

4 Examples with Normal State-space Model

The normal state-space model is used for demonstrating the proposed technique of the prior knowledge extraction. For more simplified presentation the model with single output is used. The model is given by

$$\begin{aligned} y_t &= Du_t + Cx_t + e_t, \\ x_t &= Bu_t + Ax_{t-1} + \omega_t, \end{aligned} \tag{18}$$

where e_t is a measurement Gaussian noise with zero mean and the variance r; ω_t is a process Gaussian noise with zero mean values and covariance Q; A, B, C and D are the matrices of known model parameters with appropriate dimensions.

4.1 Example for $\tau = 0$

With the zero pair $f(y_0, u_0)$ of expert knowledge pieces available the following sequence of calculations is necessary. As it has been noted in Section 3, one can use here the normal observation model only. It is given by

$$\begin{aligned} & f\left(y_0 | u_0, x_0\right) \qquad (19) \\ &= \underbrace{(2\pi r)^{-0.5}}_{A(x_0)} \exp\left\{-\frac{1}{2r}(y_0 - Du_0 - Cx_0)^2\right\}, \\ &= A(x_0) \\ &\times \exp\left\{ \overbrace{-\frac{1}{2} tr\left(\underbrace{\frac{\xi_0'\xi_0}{r}}_{B(d^{\,0})} \underbrace{[1\;1\;x_0']'[1\;1\;x_0']}_{C(x_0)}\right)}^{<B(d^{\,0}),C(x_0)>} \right\}, \end{aligned}$$

where vector $\xi_0 = [y_0 \quad - Du_0 \quad - C]$, tr is a trace of the matrix. The calculation is straightforward by using multiplication of the matrices and properties of the inner product.

Since the variance is supposed to be known, the conjugate prior Gaussian pdf can be chosen with the mean (column) vector μ and covariance P.

$$\begin{aligned} f(x_0) &\propto (2\pi)^{-\frac{\mathring{x}}{2}}|P|^{-\frac{1}{2}} \qquad (20) \\ &\times \exp\left\{-\frac{1}{2}\left[x_0 - \mu\right]' \frac{1}{P}\left[x_0 - \mu\right]\right\}, \\ &\propto \frac{(2\pi)^{-\frac{\mathring{x}}{2}}}{\sqrt{|P|}} \\ &\times \exp\left\{-\frac{1}{2} tr\left(\bar{V}[1\;1\;x_0']'[1\;1\;x_0']\right)\right\}, \end{aligned}$$

where $\mathring{x}$ is a number of entries of the state vector (always column) x_t and

$$\bar{V} = M\frac{1}{P}M', \text{ with} \qquad (21)$$

M as a matrix, composed from two diagonal matrices, one above another. The first upper one contains the negative elements from the prior mean vector μ at the diagonal, the second lower matrix is a unit matrix of appropriate dimension.

Since the conjugate prior pdf is a Gaussian one, the transformed pdf $f\left(x_0 | \tau^*\right)$ is also Gaussian and preserves the same form. With the help of multiplication of (20) and (19) the following statistic is obtained for it.

$$V_0 = \bar{V} + \underbrace{\frac{\xi_0'\xi_0}{r}}_{B(d^{\,0})}. \qquad (22)$$

The proposed statistic defines the update of the initial state x_0, incorporating the zero index pair of the expert knowledge. By the straightforward calculation of matrix V_0, its partition as $[F_0 \;\; G_0';\;\; G_0 \;\; H_0]$, where F_0, G_0, H_0 are the square matrices, and by completion the squares for x_0, the mean and covariance of the initial state x_0, conditional on the available expert knowledge, can be obtained. The estimation is similar to Kalman filter' data updating, and the covariance matrix and the mean vector are as follows.

$$\begin{aligned} \hat{P}_0 &= H_0^{-1}, \qquad (23) \\ \hat{x}_0 &= -H_0^{-1}G_0[1\;\;1]', \qquad (24) \end{aligned}$$

where, for example, for two-dimensional state and input

$$\begin{aligned} H_0 &= \begin{bmatrix} \tilde{p}_{11} + \frac{c_1^2}{r} & \tilde{p}_{12} + \frac{c_1 c_2}{r} \\ \tilde{p}_{21} + \frac{c_2 c_1}{r} & \tilde{p}_{22} + \frac{c_2^2}{r} \end{bmatrix}, \\ G_0 &= \begin{bmatrix} -\tilde{p}_{11}\mu_1 - \frac{c_1 y_0}{r} & -\tilde{p}_{12}\mu_2 + \frac{c_1 Du}{r} \\ -\tilde{p}_{21}\mu_1 - \frac{c_2 y_0}{r} & \tilde{p}_{22}\mu_2 + \frac{c_2 Du}{r} \end{bmatrix}, \end{aligned}$$

with $\tilde{p}_{ij}$ as the elements from $\tilde{P} = P^{-1}$.

4.2 Example for $\tau = 1$

For the case, when available expert knowledge contains $f(y_0, y_1, u_0, u_1)$, the model is given by the pdf (9) and has the following form, when it is Gaussian one.

$$\begin{aligned} & f\left(y_0, y_1, u_0, u_1 | x_0\right) \\ &= \int (2\pi r_1)^{-0.5} \exp\left\{-\frac{1}{2r_1}(y_1 - Du_1 - Cx_1)^2\right\}, \\ &\times (2\pi)^{-\frac{\mathring{x}}{2}}|Q|^{-\frac{1}{2}} \\ &\times \exp\left\{-\frac{1}{2}\left[x_1 - \bar{x}_1\right]' \frac{1}{P}\left[x_1 - \bar{x}_1\right]\right\} dx_1 \\ &\times (2\pi r)^{-0.5} \\ &\times \exp\left\{-\frac{1}{2} tr\left(\frac{\xi_0'\xi_0}{r}[1\;1\;x_0']'[1\;1\;x_0']\right)\right\}, \end{aligned}$$

where r_1 is a measurement noise variance and $\bar{x}_1 = Bu_1 - Ax_0$. The latter is used only for shorthand notation here.

In order to integrate over x_1 it is necessary to rearrange the quadratic forms inside the exponents and fulfill the completion of squares for x_1. After that and applying the matrix inversion lemma [4] two quadratic form inside the integral are modified as

$$[x_1 - \hat{x}_1]' \frac{1}{\hat{Q}} [x_1 - \hat{x}_1] + \frac{1}{\hat{r}_1}(y_1 - \hat{y}_1)^2, \qquad (25)$$

where

$$\begin{aligned} \hat{Q}^{-1} &= Q^{-1} + \frac{C'C}{r_1}, \\ \hat{x}_1 &= \hat{Q}[Q^{-1}(Bu_1 + Ax_0) + \frac{C'}{r_1}(y_1 - Du_1)], \\ \hat{r}_1 &= r_1 + CQC', \\ \hat{y}_1 &= Du_1 + C(Bu_1 + Ax_0). \end{aligned}$$

After integrating the model takes the form

$$A(x_0) \exp\left\{ -\frac{1}{2r_1}(y_1 - \hat{y}_1)^2 \right\} \qquad (26)$$

$$\times \exp\left\{ -\frac{1}{2} tr \left(\frac{\xi_0'\xi_0}{r} [1\ 1\ x_0']'[1\ 1\ x_0'] \right) \right\},$$

$$= A(x_0) \exp\left\{ -\frac{1}{2} tr \left((\underbrace{\frac{\xi_1'\xi_1}{r_1}}_{B(d^{\,1})} + B(d^{\,0}))C(x_0) \right) \right\}$$

where

$$\begin{aligned} A(x_0) &= \frac{1}{2\pi}(r_1 r |Q||\hat{Q}|)^{-0.5}, \qquad (27) \\ \xi_1 &= [y_1 \quad -\tilde{D}u_0 \quad -\tilde{C}]; \\ \tilde{D} &= D + CB, \ \tilde{C} = CA. \end{aligned}$$

It is obviously, that statistic for the first and zero pair of the expert knowledge available is calculated as

$$V_1 = \underbrace{\bar{V} + B(d^{\,0})}_{V_0} + \underbrace{\frac{\xi_1'\xi_1}{r_1}}_{B(d^{\,1})}. \qquad (28)$$

The mean vector and the covariance matrix of the initial state distribution are obtained similarly to the previous step with the help of partition of matrix V_1.

4.3 Recursive Statistics For Normal Model

Thus, calculating the statistics V with a subsequent index τ, similarly as in two previous steps, demonstrated in Sections 4.1 and 4.2, one can obtain the following recursively calculated relations.

$$\begin{aligned} V_\tau &= V_{\tau-1} + \underbrace{\frac{\xi_\tau'\xi_\tau}{r_\tau}}_{B(d^{\,\tau})}, \qquad (29) \\ V_\tau &= \begin{bmatrix} F_\tau & G_\tau' \\ G_\tau & H_\tau \end{bmatrix}, \\ \hat{P}_0(\tau) &= H_\tau^{-1}, \\ \hat{x}_0(\tau) &= -H_\tau^{-1} G_\tau [1\ \ 1]', \end{aligned}$$

5 Conclusion

The paper describes the approach to prior knowledge quantification, applied for state estimation. The application of the presented technique provides the initial state distribution, conditional on available expert information. The resulting distribution is used as the initial conditions for Kalman filtering.

The application area of the paper is the traffic control, where the expert knowledge are often available as the simulated, or historical data. The experiments with the incorporation of the simulated data for the choice of the initial state distribution are the next task to be solved.

Acknowledgements: The research was supported by GA ČR grant No. 201/06/P434, AV ČR project BADDYR No. 1ET100750401 and grant of Ministry of Transportation of the Czech Republic No. 1F43A/003/120.

References:

[1] Greg Welch and Gary Bishop, "An Introduction to the Kalman Filter", Tech. Rep. 95-041, UNC-CH Computer Science, 1995.

[2] M. Kárný, P. Nedoma, N. Khailova, and L. Pavelková, "Prior information in structure estimation", *IEE Proceedings — Control Theory and Applications*, vol. 150, no. 6, pp. 643–653, 2003.

[3] Kárný, J. Andrýsek, A. Bodini, T.V. Guy, J. Kracík, P. Nedoma, and F. Ruggeri, "Fully probabilistic knowledge expression and incorporation", *IEEE Trans. on Automatic Control*, 2006, submitted.

[4] V. Peterka, "Bayesian system identification", in *Trends and Progress in System Identification*, P. Eykhoff, Ed., pp. 239–304. Pergamon Press, Oxford, 1981.