Akademie věd České republiky
Ústav teorie informace a automatizace

Academy of Sciences of the Czech Republic
Institute of Information Theory and Automation

RESEARCH REPORT

E. Suzdaleva, M. Kárný

Factorized Filtering

2161 June 2006

GAČR 201/06/P434, AVČR 1ET100750401

ÚTIA AVČR, P.O.Box 18, 182 08 Prague,
Czech Republic
Fax: (+420)286890378, http://www.utia.cas.cz, E-mail:
utia@utia.cas.cz

Contents

Chapter 1

Introduction

1.1 Motivation

Dynamic decision making under uncertainty forms the core of such seemingly diverse activities as parameter estimation [1], filtering [2], testing of hypothesis [3], pattern recognition [4], advising [5], fault detection [6], feedback control [7, 8] etc. Practical needs and increasing computing power make designers to address problems of permanently increasing complexity. This successful trend reveals inherent limitations of the available tools and stimulates attempts to overcome them.

One of the potential application of this work is the urban traffic control systems. The fact, that the traffic control tools are significantly limited, surprises nobody: modern, powerful cars have to move slowly and inefficiently through towns within permanently extending peak hours. Adequate extension of traffic network is expensive and often impossible, especially in historical towns. Thus, it calls for exploiting all available means starting from economical pressure, various regulative measures up to exploitation of modern, ideally adaptive, feedback control.

1.2 Addressed Problem and Layout

One of the main controlled variables in traffic systems is a queue length, which expresses the optimality of a traffic network most adequately. It is directly unobserved and, therefore, has to be estimated. At the same time, other state variables are of a discrete-valued nature. Thus, estimation of mixed-type data (continuous and discrete valued) models is highly desirable. A potential solution to this problem calls for a factorized version of the state estimation, which allows to model the entries of the state individually. In this way, the task *how to obtain the estimates of the individual time-varying state factors* is addressed in the paper.

The layout of the paper is the following one. The subsequent section is devoted to the state of the art of the problem. The Chapter 2 provides the necessary basic facts, including the notations, used throughout the text, the models and describes the general form of the prediction and filtering. The Chapter 3 offers the factorized version of the prediction and filtering and demonstrates the example with Gaussian single output system. The remarks close the paper.

1.3 State of the Art

In Bayesian methodology, adopted in the work, the factorized version of the filter is obtained by applying chain rule to the state-space model. The problem was already solved for a degenerate case of time-invariant state, which coincides with parameter estimation [5]. The obtained results indicated a chance to update posterior probability density functions in the entry-wise manner.

The state of the art of the problem includes a series of research in the field. Most works found are devoted to factorization of well-known Kalman filter [9]. Despite the variety of the research at this area, the global aim of majority of these works is reduction of the computational complexity with the help of lesser rank of the covariance matrix, but not the obtaining of the estimates of the individual state entries, which is the aim of the present work. For example, the work [10] deals with factorization of the covariance matrix in Kalman filter, where the covariance matrix was decomposed with the help of square root factorization. The QR-factorized filter and smoother algorithms for use on linear time-varying discrete-time problems, that can handle the general case of a singular state transition matrix, are discussed in [11]. The UD-factorization of Kalman filter for the multi-sensor data fusion is presented in [12]. Another work, devoted to the UD-factorized covariance filter application, is concerned with development of a connected element interferometer [13]. The method for particle filtering, which factorizes the likelihood, was proposed in [14]. It considers the problem, when the state space can be partitioned in groups of random variables, whose likelihood can be independently evaluated.

As regards the nonlinear estimation, the following research works should be noted here. The square root form of unscented Kalman filter (UKF) for the state and parameter estimation, which, in its turn, was proposed as an alternative to the extended Kalman filter, used for nonlinear estimation, is described in [15]. This square-root UKF has better numerical properties and guarantees positive semi-definiteness of the underlying state covariance.

The factorization of the covariance matrices is also used in problems of systems classification, dealing with multivariate Gaussian random field [16].

The problem of filtering with Gaussian models can be also considered with the help of dynamic Bayesian networks [17]. Within this framework the problem of joint state-parameter estimation is often met.

The work [18] was directed exactly at the estimation of the individual state factors, and it proposed the recursive algorithm of factorized filtering, requiring a special, reduced, form of the state-space model. The present paper offers the solution without such a restriction.

Moreover, the overview of the problem showed the results mostly with Gaussian models. The general solution, based on the Bayesian framework, can be potentially helpful in the case of other models.

Chapter 2

Preliminaries

2.1 General Conventions

The notations listed here are mostly followed in this work. If some exception is necessary it is introduced at the place of its validity. If some verbal notions are introduced within Propositions, Remarks, etc., then they are *emphasized*.

$\equiv$ *means the equality by definition.*

$f(|)$ *is the letter reserved for conditional probability (density) functions (p(d)f).*

The meaning of the p(d)f is given through the name of its argument. When the argument x coincides with realization of the corresponding random variable then it is made bold, i.e. $f(\boldsymbol{x})$.

x^* *denotes the range of* x, $x \in x^*$.

$\mathring{x}$ *denotes the number of members in the countable set* x^* or the *number of entries in the vector* x.

x_t *is a quantity* x *at the* discrete time *instant labelled by* $t \in t^* \equiv \{1, \ldots, \mathring{t}\}$.

$\mathring{t} \leq \infty$ *is called (decision, learning, prediction, control, advising) horizon.*

x_t^i *is an ith entry of the array* x *at time* t.

The subscript symbol is a time index.

$x^{k:l}$ *denotes the sequence* $x_k, \ldots, x_l$.

$x^{l:k}$ is an empty sequence and reflects just the prior information if $l < k$.

$\mathrm{supp}\,[f(x)]$ *is the support of the pdf* $f() : x^* \to [0, \infty]$, *i.e., the subset of* x^* *on which* $f(x) > 0$.

2.2 Models

Here, general basis of Bayesian learning is summarized.

Agreement 2.2.1 (Models in DM)
The model of observation

$$f\left(y_t \middle| u_t, d^{1:t-1}, x_t\right), \; t \in t^*, \tag{2.1}$$

relates outputs y_t *to the current actions* u_t*, the past data* $d^{1:t-1}$*, and current state* x_t*.*
The model of evolution

$$f\left(x_t \middle| u_t, d^{1:t-1}, x_{t-1}\right), \; t \in t^*, \tag{2.2}$$

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

$$f\left(u_t \middle| d^{1:t-1}, x_t\right), \; t \in t^*, \tag{2.3}$$

describes, generally randomized, generating of actions u_t *based on* $d^{1:t-1}, x_t$*.*

By definition, admissible strategies cannot exploit directly unobserved state, i.e., they have to meet so called natural conditions of DM *(*NCDM*)*

$$f\left(u_t \middle| d^{1:t-1}, x_t\right) = f\left(u_t \middle| d^{1:t-1}\right), \; t \in t^*. \tag{2.4}$$

2.3 Prediction and Filtering

Proposition 2.3.1 (Prediction and filtering) *Under natural conditions of control, the* predictor $f\left(y_t \middle| u_t, d^{1:t-1}\right)$ *is given by the formula*

$$f\left(y_t \middle| u_t, d^{1:t-1}\right) = \int f\left(y_t \middle| u_t, d^{1:t-1}, x_t\right) f\left(x_t \middle| u_t, d^{1:t-1}\right) \, dx_t. \tag{2.5}$$

The pdf $f\left(x_t \middle| u_t, d^{1:t-1}\right)$ *estimating the state* x_t *evolves according the following coupled formulas.*
Data updating

$$\begin{aligned} f\left(x_t \middle| d^{1:t}\right) &= \frac{f\left(y_t \middle| u_t, d^{1:t-1}, x_t\right) f\left(x_t \middle| u_t, d^{1:t-1}\right)}{f\left(y_t \middle| u_t, d^{1:t-1}\right)} \\ &\propto f\left(y_t \middle| u_t, d^{1:t-1}, x_t\right) f\left(x_t \middle| u_t, d^{1:t-1}\right) \end{aligned} \tag{2.6}$$

that incorporates the experience contained in the data record d_t *consisting of the output* y_t *and the input* u_t*.*
Time updating

$$f\left(x_{t+1} \middle| u_{t+1}, d^{1:t}\right) \propto \int f\left(x_{t+1} \middle| u_{t+1}, d^{1:t}, x_t\right) f\left(x_t \middle| d^{1:t}\right) \, dx_t. \tag{2.7}$$

The recursions start from the prior pdf $f\left(x_1 \middle| u_1\right)$ *that expresses the subjective prior knowledge on the initial state* x_1*.*

Proof: Implied by marginalization and the chain rule. □

Chapter 3

Factorized Filtering

3.1 Factorized Version of Prediction and Filtering

Proposition 3.1.1 (Factorized prediction and filtering) *Under natural conditions of control, the predictor* $f\left(y_t \middle| u_t, d^{1:t-1}\right)$ *is the last pdf of the sequence of the partially conditioned predictors*

$$f\left(y_t \middle| u_t, d^{1:t-1}, x_t^{i+1:\mathring{x}}\right) = \int f\left(y_t \middle| u_t, d^{1:t-1}, x_t^{i:\mathring{x}}\right) f\left(x_t^i \middle| u_t, d^{1:t-1}, x_t^{i+1:\mathring{x}}\right) dx_t^i. \tag{3.1}$$

The pdfs $f\left(x_t^i \middle| u_t, d^{1:t-1}, x_t^{i+1:\mathring{x}}\right)$, $i = 1, \ldots, \mathring{x}$, *determining the state estimate through the chain rule* $f\left(x_t \middle| u_t, d^{1:t-1}\right) = \prod_{i=1}^{\mathring{x}} f\left(x_t^i \middle| u_t, d^{1:t-1}, x_t^{i+1:\mathring{x}}\right)$ *evolve according the following coupled formulas.*
Data updating

$$\begin{aligned} f\left(x_t^i \middle| d^{1:t}, x_t^{i+1:\mathring{x}}\right) &= \frac{f\left(y_t \middle| u_t, d^{1:t-1}, x_t^{i:\mathring{x}}\right) f\left(x_t^i \middle| u_t, d^{1:t-1}, x_t^{i+1:\mathring{x}}\right)}{f\left(y_t \middle| u_t, d^{1:t-1}, x_t^{i+1:\mathring{x}}\right)} \\ &\propto f\left(y_t \middle| u_t, d^{1:t-1}, x_t^{i:\mathring{x}}\right) f\left(x_t^i \middle| u_t, d^{1:t-1}, x_t^{i+1:\mathring{x}}\right) \end{aligned} \tag{3.2}$$

that incorporates the experience contained in the data record d_t *consisting of the output* y_t *and the input* u_t*.*
Time updating
The pdf $f\left(x_{t+1} \middle| u_{t+1}, d^{1:t}\right)$ *is the last member of the sequence of the partially conditioned state estimates indexed by* $j = 1, \ldots, \mathring{x}$

$$\begin{aligned} & f\left(x_{t+1} \middle| u_{t+1}, d^{1:t}, x_t^{j+1:\mathring{x}}\right) \\ \propto & \int f\left(x_{t+1} \middle| u_{t+1}, d^{1:t}, x_t^{j:\mathring{x}}\right) f\left(x_t^j \middle| d^{1:t}, x_t^{j+1:\mathring{x}}\right) dx_t^j. \end{aligned} \tag{3.3}$$

The recursions start from the prior pdfs $f\left(x_1 \middle| u_1\right)$ *that expresses the subjective prior knowledge on the initial sequence of internal quantities.*

The factorized version of the state estimate after time-updating is obtained by straightforward application of the chain rule for pdfs.

Proof: Evaluations follows Proposition 2.3.1 and use Fubini theorem on multiple integration [19]. Let us demonstrate its application on evaluation of the predictor.

$$\begin{aligned} & f\left(y_t \middle| u_t, d^{1:t-1}\right) = \int f\left(y_t \middle| u_t, d^{1:t-1}, x_t\right) f\left(x_t \middle| u_t, d^{1:t-1}\right) dx_t \\ = & \underbrace{\int \ldots \int \ldots \underbrace{\int f\left(y_t \middle| u_t, d^{1:t-1}, x_t\right) f\left(x_t^1 \middle| u_t, d^{1:t-1}, x_t^{2:\mathring{x}}\right) dx_t^1}_{f\left(y_t \middle| u_t, d^{1:t-1}, x_t^{2:\mathring{x}}\right)} \ldots f\left(x_t^i \middle| u_t, d^{1:t-1}, x_t^{i+1:\mathring{x}}\right) dx_t^i}_{f\left(y_t \middle| u_t, d^{1:t-1}, x_t^{i+1:\mathring{x}}\right)} \\ \times & f\left(x_t^{i+1} \middle| u_t, d^{1:t-1}, x_t^{i+2:\mathring{x}}\right) \ldots \times f\left(x_t^{\mathring{x}} \middle| u_t, d^{1:t-1}\right) dx_t^{\mathring{x}} \end{aligned}$$

Data updating (3.2) is just application of Bayes rule. Time updating of pdfs (3.3) conditioned by entries of the older states is direct application of the general time-updating. Similarly as for prediction, jth evaluation provides the needed transition kernel for next evaluation. □

Remark 3.1.1

1. *The proposed factorization can be made even more extensive by factorizing the observation model. This line will not be followed here to simplify the presentation. But practical algorithms should deal with such a factorization, too.*

3.2 Factorized Kalman filtering for single output system

Here, we apply Proposition 3.1.1 to linear Gaussian state space model with Gaussian prior on initial state. This state estimation is known to be described by Kalman filter [2].

Throughout, we use repeatedly the following simple proposition.

Proposition 3.2.1 (Square completion and integration of Gaussian pdf) *For real scalars x, α, β, γ and positive scalars r, p, it holds*

$$
\begin{aligned}
h(x) &\equiv \exp\left\{-\frac{(\beta-\gamma x)^2}{2r} - \frac{(x-\alpha)^2}{2p}\right\} = \exp\left\{-\frac{(x-\hat{x})^2}{2R} - \frac{\lambda}{2}\right\} \text{ with} \qquad (3.4)\\
R &= \frac{rp}{r+\gamma^2 p}, \quad \hat{x} = \frac{\alpha r + \beta\gamma p}{r+\gamma^2 p}, \quad \lambda = \frac{(\beta-\alpha\gamma)^2}{r+\gamma^2 p}\\
\int h(x)\,dx &= \sqrt{2\pi R}\exp\left\{-\frac{\lambda}{2}\right\}.
\end{aligned}
$$

Proof: By a direct comparison of quadratic forms in x in both exponents and by noticing that x-dependent factor of $h(x)$ is non-normalized Gaussian pdf. □

The state estimate is assumed in the form

$$
f\left(x_t^i \middle| u_t, d^{1:t-1}, x_t^{i+1:\mathring{x}}\right) = \mathcal{N}_{x_t^i}\left(\hat{\mu}_{t|t-1;i} + \sum_{k=i+1}^{\mathring{x}} g_{t|t-1;ik} x_{t;k}, p_{t|t-1;i}\right), \qquad (3.5)
$$

where $\hat{\mu}_{t|t-1;i}$ is the term depending on the data $u_t, d^{1:t-1}$ only, $g_{t|t-1;ik}$ are coefficients, which are data and state independent similarly as the variance $p_{t|t-1;i} > 0$.

For presentation simplicity, single output case, $\mathring{y}$ is considered. This helps us to avoid consequences of the incomplete factorization, cf. Remark 3.1.1. Thus, we assume

$$
f\left(y_t \middle| u_t, d^{1:t-1}, x_t\right) = \mathcal{N}_{y_t}\left(\rho_t + \sum_{k=1}^{\mathring{x}} c_{t;k} x_{t;k}, r_t\right), \qquad (3.6)
$$

where the offset ρ_t, coefficients $c_{t;k}$, $k = 1, \ldots, \mathring{x}$, and variance r_t are assumed to be known functions of $u_t, d^{1:t-1}$.

Proposition 3.2.2 (Partially conditioned Gaussian observation models) *For the Gaussian factors of the state estimate (3.5) and the observation model (3.6), it holds*

$$
f\left(y_t \middle| u_t, d^{1:t-1}, x_t^{i:\mathring{x}}\right) = \mathcal{N}_{y_t}\left(\rho_{t;i} + \sum_{k=i}^{\mathring{x}} c_{t;ik} x_{t;k}, r_{t;i}\right), \qquad (3.7)
$$

where the state independent offsets $\rho_{t;i}$, coefficients $c_{t;ik}$, $k = i, \ldots, \mathring{x}$, evolve according to the following recursions, for $i = 1, \ldots, \mathring{x}$

$$
\begin{aligned}
\rho_{t;i+1} &= \rho_{t;i} + c_{t;ii}\hat{\mu}_{t|t-1;i} \qquad (3.8)\\
c_{t;(i+1)k} &= c_{t;ik} + c_{t;ii} g_{t|t-1;ik}, \text{ for } k > i\\
r_{t;i+1} &= r_{t;i} + p_{t|t-1;i} c_{t;ii}^2.
\end{aligned}
$$

The recursions start from $\rho_{t;1} = \rho_t$, $c_{t;1k} = c_{t;k}$ and $r_{t;1} = r_t$.

Proof: For $i = 1$, the assumed form coincides with the observation model (3.6). This verifies the first inductive step and determines initial conditions of the recursion. In the ith induction step, we use identity (3.1), the state estimate (3.5) and the observation model (3.6)

$$
\begin{aligned}
f\left(y_t \middle| u_t, d^{1:t-1}, x_t^{i+1:\mathring{x}}\right) &= \int f\left(y_t \middle| u_t, d^{1:t-1}, x_t^{i:\mathring{x}}\right) f\left(x_t^i \middle| u_t, d^{1:t-1}, x_t^{i+1:\mathring{x}}\right) dx_t^i\\
\propto \int \exp &\left\{-\frac{\left(y_t - \rho_{t;i} - \sum_{k=i}^{\mathring{x}} c_{t;ik} x_{t;k}\right)^2}{2r_{t;i}} - \frac{\left(x_{t;i} - \hat{\mu}_{t|t-1;i} - \sum_{k=i+1}^{\mathring{x}} g_{t|t-1;ik} x_{t;k}\right)^2}{2p_{t|t-1;i}}\right\} dx_t^i.
\end{aligned}
$$

The integrand has the form (3.4) with the correspondence

$$\begin{aligned}\alpha &\equiv \hat{\mu}_{t|t-1;i} + \sum_{k=i+1}^{\mathring{x}} g_{t|t-1;ik} x_{t;k}, \quad p \equiv p_{t|t-1;i}, \quad r \equiv r_{t;i} \\ \beta &\equiv y_t - \rho_{t;i} - \sum_{k=i+1}^{\mathring{x}} c_{t;ik} x_{t;k}, \quad \gamma \equiv c_{t;ii}, \quad x \equiv x_{t;i}.\end{aligned}$$

Proposition 3.2.1 gives the value of integral in the Gaussian form with the corresponding λ in exponent. The claimed recursions are then implied by a trivial re-arrangement of arguments in it. □

Remark 3.2.1

1. *The recursions contain no numerically dangerous operation.*

Proposition 3.2.3 (Factorized data updating) *The functional form of the state estimate (3.5) preserves in data updating. Specifically, after data updating*

$$f\left(x_t^i \middle| d^{1:t}, x_t^{i+1:\mathring{x}}\right) = \mathcal{N}_{x_{t;i}}\left(\hat{\mu}_{t|t;i} + \sum_{k=i+1}^{\mathring{x}} g_{t|t;ik} x_{t;k}, p_{t|t;i}\right), \tag{3.9}$$

with

$$\begin{aligned}K_{t|t;i} &\equiv \frac{c_{t;ii} p_{t;i}}{r_{t;i+1}} \\ \hat{\mu}_{t|t;i} &= \hat{\mu}_{t|t-1;i} + K_{t|t;i}(y_t - \rho_{t;i} - c_{t;ii}\hat{\mu}_{t|t-1;i}) \\ g_{t|t;ik} &= g_{t|t-1;ik} - K_{t|t;i}(c_{t;ik} + c_{t;ii} g_{t|t-1;ik}) \text{ for } k > i \\ p_{t|t;i} &= \frac{r_{t;i}}{r_{t;i+1}} p_{t|t-1;i}\end{aligned} \tag{3.10}$$

Proof: Data updating essentially completes evaluations used in proof of Proposition 3.2.2. Proposition 3.1.1, (3.2) applied to the partial observation models (3.8) and assumed form of the state estimate (3.5) implies

$$\begin{aligned} f\left(x_t^i \middle| d^{1:t}, x_t^{i+1:\mathring{x}}\right) &= \frac{f\left(y_t \middle| u_t, d^{1:t-1}, x_t^{i:\mathring{x}}\right) f\left(x_t^i \middle| u_t, d^{1:t-1}, x_t^{i+1:\mathring{x}}\right)}{f\left(y_t \middle| u_t, d^{1:t-1}, x_t^{i+1:\mathring{x}}\right)} \\ \underbrace{\propto}_{(3.7)} \exp&\left\{-\frac{\left(y_t - \rho_{t;i} - \sum_{k=i}^{\mathring{x}} c_{t;ik} x_{t;k}\right)^2}{r_{t;i}} - \frac{\left(x_{t;i} - \hat{\mu}_{t|t-1;i} - \sum_{k=i+1}^{\mathring{x}} g_{t|t-1;ik} x_{t;k}\right)^2}{p_{t|t-1;i}}\right\}. \end{aligned}$$

The application of Proposition 3.2.1 with the correspondence

$$\begin{aligned}\alpha &\equiv \hat{\mu}_{t|t-1;i} + \sum_{k=i+1}^{\mathring{x}} g_{t|t-1;ik} x_{t;k}, \quad p \equiv p_{t|t-1;i}, \quad r \equiv r_{t;i} \\ \beta &\equiv y_t - \rho_{t;i} - \sum_{k=i+1}^{\mathring{x}} c_{t;ik} x_{t;k}, \quad \gamma \equiv c_{t;ii}, \quad x \equiv x_{t;i}.\end{aligned}$$

gives the Gaussian form. The rest is again based on re-arrangement and definition of $r_{t;i+1}$ (3.8). In the re-arrangement based terms $\hat{\mu}_{t|t-1;i}$, $g_{t|t-1;i}$ are added and subtracted, respectively, in order to get usual forms of recursions. □

For time updating, we have to evaluate partially conditioned linear Gaussian time-evolution model. The chain rule implies that the fully conditioned model can be always given the form

$$\begin{aligned} f\left(x_{t+1} \middle| u_{t+1}, d^{1:t}, x_t\right) &= \prod_{i=1}^{\mathring{x}} f\left(x_{t+1}^i \middle| x_{t+1}^{i+1:\mathring{x}}, u_{t+1}, d^{1:t}, x_t\right) \\ &= \prod_{i=1}^{\mathring{x}} \mathcal{N}_{x_{t+1}^i}\left(\zeta_{t+1;i} + \sum_{k=i+1}^{\mathring{x}} \alpha_{t+1;ik} x_{t+1;ik} + \sum_{k=1}^{\mathring{x}} \beta_{t+1;ik} x_{t;k}, R_{t+1;i}\right), \end{aligned} \tag{3.11}$$

where for all $i \in \{1, \ldots, \mathring{x}\}$ the offset $\zeta_{t+1;i}$, coefficients $\alpha_{t+1;ik}$, $k = i+1, \ldots, \mathring{x}$, $\beta_{t+1;ik}$, $k = 1, \ldots, \mathring{x}$ and variances $R_{t+1;i}$ are assumed to be known functions of $u_{t+1}, d^{1:t}$.

The scalar variable x_t^j to be integrated out occurs now in all factors in (3.11). Thus, we need a "vector" version of Proposition 3.2.1.

Proposition 3.2.4 (Integration of a product of Gaussian pdfs) *For real scalar x, vectors $\beta \equiv [\beta_1, \ldots, \beta_{\mathring{\beta}}]'$, $\gamma \equiv [\gamma_1, \ldots, \gamma_{\mathring{\beta}}]'$ and diagonal precision matrix $\omega \equiv diag\left[r_1^{-1}, \ldots, r_{\mathring{\beta}}^{-1}\right]$, it holds*

$$\int h(x)\, dx \equiv \int \exp\left\{-\sum_{i=1}^{\mathring{\beta}} \frac{(\beta_i - \gamma_i x)^2}{2r_i}\right\} dx \propto \exp\left\{-\frac{\lambda}{2}\right\} \text{ with} \tag{3.12}$$

$$\lambda \equiv \beta'\left(\omega - \frac{\omega\gamma\gamma'\omega}{\gamma'\omega\gamma}\right)\beta \equiv \sum_{l=1}^{\mathring{\beta}-1} \frac{\left(\sum_{i=l}^{\mathring{\beta}} U_{li}\beta_i\right)^2}{p_l}, \text{ where}$$

the upper triangular $(\mathring{\beta}-1, \mathring{\beta})$ matrix U with unit diagonal is found via $U'DU$ decomposition

$$U'\text{diag}\left[p_1^{-1}, \ldots, p_{\mathring{\beta}}^{-1}\right] U = \omega - \frac{\omega\gamma\gamma'\omega}{\gamma'\omega\gamma}. \tag{3.13}$$

Proof: By completion of squares with respect to the scalar x and integration of univariate Gaussian pdf. The generic rank-one deficit follows from the projector type form of right-hand side in (3.13). □

Remark 3.2.2

1. *The mentioned matrix decomposition can be performed similarly to the algorithm REFIL [1].*

Proposition 3.2.5 (Partially conditioned Gaussian time-evolution models, Factorized time-updating) *For the Gaussian factors of the state estimate (3.9) and the time-evolution model (3.11), it holds*

$$f\left(x_{t+1;i}\middle| x_{t+1}^{i+1:\mathring{x}}, u_{t+1}, d^{1:t}, x_t^{j+1:\mathring{x}}\right) \tag{3.14}$$
$$= \mathcal{N}_{x_{t+1;i}}\left(\hat{\mu}_{t+1|t;ij} + \sum_{k=i+1}^{\mathring{x}} g_{t+1|t;ikj} x_{t+1;k} + \sum_{k=j+1}^{\mathring{x}} \beta_{ikj} x_{t;k}, p_{t+1|t;ij}\right),$$

where for $i \in \{1, \ldots, \mathring{x}\}$ offsets $\hat{\mu}_{t+1|t;ij}$, coefficients $g_{t+1|t;ikj}$, $k = i+1, \ldots, \mathring{x}$, β_{ikj}, $k = j+1, \ldots, \mathring{x}$ and variances $p_{t+1|t;ij} > 0$ are the state independent. The following recursions over $j = 1, \ldots, \mathring{x}$ hold

$$\hat{\mu}_{t+1|t;i(j+1)} = \hat{\mu}_{t+1|t;ij} + \sum_{l=i+1}^{\mathring{x}} U_{il}\hat{\mu}_{t+1|t;lj} - U_{i(\mathring{x}+1)}\hat{\mu}_{t|t;j} \tag{3.15}$$

$$g_{t+1|t;ik(j+1)} = g_{t+1|t;ikj} + \sum_{l=i-1}^{\mathring{x}} U_{il} g_{t+1|t;lkj} - \sum_{l=i+1}^{\mathring{x}} U_{il}$$

$$\beta_{ik(j+1)} = \beta_{ikj} + \sum_{l=i+1}^{\mathring{x}} U_{il}\beta_{lkj} - U_{i(\mathring{x}+1)} g_{t|t;jk},\ k > j. \tag{3.16}$$

The upper triangular $(\mathring{x}, \mathring{x}+1)$ matrix U with unit diagonal as well as the positive scalars $p_{t+1|t;i(j+1)}$ are obtained via the decomposition

$$\omega - \frac{\omega\gamma\gamma'\omega}{\gamma'\omega\gamma} \equiv U'\text{diag}\left[p_{t+1|t;1(j+1)}^{-1}, \ldots, p_{t+1|t;\mathring{x}(j+1)}^{-1}\right] U \text{ with} \tag{3.17}$$
$$\gamma' \equiv [\beta_{1jj}, \ldots, \beta_{\mathring{x}jj}, 1]$$
$$\omega \equiv \text{diag}\left[p_{t+1|t;1j}^{-1}, \ldots, p_{t+1|t;\mathring{x}j}^{-1}, p_{t|tj}^{-1}\right].$$

The recursions start from $\hat{\mu}_{t+1|t;i1} = \zeta_{t+1;i}$, and $p_{t+1|t;i1} = R_{t+1;i}$, $g_{t+1|t;ik1} = \alpha_{t+1;ik}$ and $\beta_{ik1} = \beta_{t+1;ik}$.

The result obtained after the step $\mathring{x}$ provides parameters of the time-updated factors.

Proof: For $j = 1$, the partially conditioned time-evolution model coincides with the time-evolution model (3.11). Its factorized form determines the initial conditions of the recursion over j. This verifies the first inductive step and determines initial conditions of the recursion.

In the jth induction step, we use identity (3.3), the updated state estimate (3.9) and the time-evolution model (3.11). The symbols introduced via under-bracing refer to Proposition 3.2.4, $\mathring{\beta} = \mathring{x} + 1$.

$$f\left(x_{t+1}\middle| u_{t+1}, d^{1:t}, x_t^{j+1:\mathring{x}}\right)$$

$$\propto \int \prod_{i=1}^{\mathring{x}} \exp\left\{ - \frac{\left(\underbrace{x_{t+1;i} - \hat{\mu}_{t+1|t;ij} - \sum_{k=i+1}^{\mathring{x}} g_{t+1|t;ikj} x_{t+1;k} - \sum_{k=j+1}^{\mathring{x}} \beta_{ikj} x_{t;k}}_{\beta_i} - \underbrace{\beta_{ijj}}_{\gamma_i} x_{t;j} \right)^2}{2 \underbrace{p_{t+1|t;ij}}_{r_i}} \right\}$$

$$\times \quad \exp\left\{ - \frac{\left(\underbrace{\left(\hat{\mu}_{t|t;j} + \sum_{k=j+1}^{\mathring{x}} g_{t|t;jk} x_{t;k} \right)}_{\beta_{\mathring{x}+1}} - \underbrace{1}_{\gamma_{\mathring{x}+1}} \times \underbrace{x_{t;j}}_{x} \right)^2}{2 \underbrace{p_{t|t;j}}_{r_{\mathring{x}+1}}} \right\} dx_{t;j} \propto \exp\left\{ -\frac{\lambda}{2} \right\}, \text{ where } \lambda = \sum_{i=1}^{\mathring{x}}$$

$$\frac{\left[\sum_{l=i}^{\mathring{x}} U_{il} \left(x_{t+1;l} - \hat{\mu}_{t+1|t;lj} - \sum_{k=l+1}^{\mathring{x}} g_{t+1|t;lkj} x_{t+1;k} - \sum_{k=j+1}^{\mathring{x}} \beta_{lkj} x_{t;k} \right) + U_{i(\mathring{x}+1)} \left(\hat{\mu}_{t|t;j} + \sum_{k=j+1}^{\mathring{x}} g_{t|t;jk} x_{t;k} \right) \right]^2}{p_{t+1|t;i(j+1)}}$$

$$\underbrace{=}_{U_{ii}=1} \frac{\left[x_{t+1;i} - \hat{\mu}_{t+1|t;ij} - \sum_{k=i+1}^{\mathring{x}} g_{t+1|t;ikj} x_{t+1;k} - \sum_{k=j+1}^{\mathring{x}} \beta_{ikj} x_{t;k} \right.}{}$$

$$\frac{\left. + \sum_{l=i+1}^{\mathring{x}} U_{il} \left(x_{t+1;l} - \hat{\mu}_{t+1|t;lj} - \sum_{k=l+1}^{\mathring{x}} g_{t+1|t;lkj} x_{t+1;k} - \sum_{k=j+1}^{\mathring{x}} \beta_{lkj} x_{t;k} \right) + U_{i,\mathring{x}+1} \left(\hat{\mu}_{t|t;j} + \sum_{k=j+1}^{\mathring{x}} g_{t|t;jk} x_{t;k} \right) \right]^2}{p_{t+1|t;i(j+1)}}.$$

Grouping the "constants and coefficients at $x_{t+1;k}$ and $x_{t;k}$ gives the claimed recursions. We have also used that the overall expression is product of i-dependent quadratic forms. Here, the use of the adopted "U"-decomposition pays back.

The last statement follows trivially from emptiness of the sum over the state entries with the time index t. □

3.3 Conclusion

The work proposes solution to the factorized filtering, obtained by applying the chain rule to the state-space models. Factorized filter provides the update of posterior probability density functions for the individual state entries, that can be helpful for solution to the task of the joint modelling of the mixed-type data. The recursions for calculating the factorized data updating and time updating are offered. The application of the solution is shown at the example of the linear Gaussian single output system, which gives the factorized Kalman filtering. For the factorized Kalman filtering the operations of completion of square and integration of non-normalized Gaussian probability density functions are used.

Among the advantages of the proposed approach one may note, unlike the previous solution of the factorized filtering, offered in [18], the present work does not require any restrictions for the state-space models. It should be also noted, that the proposed recursions do not contain any numerically dangerous operations.

Acknowledgements

This work was supported by GA ČR grant No. 201/06/P434 and AV ČR project BADDYR No. 1ET100750401.

Bibliography

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

[2] A.M. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.

[3] A. Wald, *Statistical Decision Functions*, John Wiley, New York, London, 1950.

[4] B.D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, London, 1997.

[5] M. Kárný, J. Böhm, T. V. Guy, L. Jirsa, I. Nagy, P. Nedoma, and L. Tesař, *Optimized Bayesian Dynamic Advising: Theory and Algorithms*, Springer, London, 2005.

[6] R. Patton, P. Frank, and R. Clark, *Fault Diagnosis in Dynamic Systems: Theory & Applications*, Prentice Hall, 1989.

[7] M.H. DeGroot, *Optimal Statistical Decisions*, McGraw-Hill, New York, 1970.

[8] H. Kushner, *Introduction to Stochastic Control*, Holt, Rinehart and Winston, New York, 1971.

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

[10] G. Dimitriu, "Implementation issues related to data assimilation using Kalman filtering", in *Proceedings for 3th World Conference on Computational Statistics & Data Analysis*, Limassol, Cyprus, October 28-31 2005, International Association for Statistical Computing.

[11] M. L. Psiaki, "Square-root information filtering and fixed-interval smoothing with singularities", *Automatica*, vol. 35, no. 7, pp. 1323–1331, July 1999.

[12] G. Girija, J. R. Raol, R. Appavu Raj, and S. Kashyap, "Tracking filter and multi-sensor data fusion", in *Sadhana. Special Issue on Advances in Modelling, System Identification & Parameter Estimation*, vol. 25, pp. 159–167. Indian Academy of Sciences, April 2000.

[13] D. Morrison, S. Pogorelc, T. Celano, and A. Gifford, "Ephemeris determination using a connected element interferometer", in *34th Annual Precise Time and Time Interval (PTTI) Meeting*, Reston, Virginia, USA, December 3-5 2002.

[14] I. Patras and M. Pantic, "Particle filtering with factorized likelihoods for tracking facial features", in *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition (FGR04)*, Seoul, Korea, May 17-19 2004.

[15] R. van der Merwe and E. A. Wan, "The square-root unscented Kalman filter for state and parameter-estimation", in *International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, Utah, May 2001.

[16] J. Saltyte-Benth and K. Ducinskas, "Linear discriminant analysis of multivariate spatial-temporal regressions", Tech. Rep. 2, Department of Mathematics, University of Oslo, July 2003, ISBN 82-553-1395-8.

[17] K. P. Murphy, "Dynamic Bayesian networks", Tech. Rep., Departments of computer science and statistics, University of British Columbia, 2002, to appear in Probabilistic Graphical Models, Michael Jordan.

[18] E. Suzdaleva, "On entry-wise organized filtering", in *Proceedings of 15th International Conference on Process Control'05*, High Tatras, Slovakia, June 7-10 2005, pp. 1–6, Department of Information Engineering and Process Control, FCFT STU, ISBN 80-227-2235-9.

[19] M.M. Rao, *Measure Theory and Integration*, John Wiley, New York, 1987.