



Faculty Publications

---

1991-02-01

## Set-valued Filtering And Smoothing

Wynn C. Stirling  
wynn\_stirling@byu.edu

Darryl Morrell

Follow this and additional works at: <https://scholarsarchive.byu.edu/facpub>



Part of the [Electrical and Computer Engineering Commons](#)

---

### BYU ScholarsArchive Citation

Stirling, Wynn C. and Morrell, Darryl, "Set-valued Filtering And Smoothing" (1991). *Faculty Publications*. 725.

<https://scholarsarchive.byu.edu/facpub/725>

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact [scholarsarchive@byu.edu](mailto:scholarsarchive@byu.edu), [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

# Set-Valued Filtering and Smoothing

Darryl R. Morrell, *Member, IEEE*, Wynn C. Stirling, *Member, IEEE*

**Abstract**—A theory of discrete-time optimal filtering and smoothing based on convex sets of probability distributions is presented. Rather than propagating a single conditional distribution as does conventional Bayesian estimation, a convex set of conditional distributions is evolved. For linear Gaussian systems, the convex set may be generated by a set of Gaussian distributions with equal covariance with means in a convex region of state space. The conventional point-valued Kalman filter is generalized to a set-valued Kalman filter, consisting of equations of evolution of a convex set of conditional means and a conditional covariance. The resulting estimator is an exact solution to the problem of running an infinity of Kalman filters and fixed-interval smoothers, each with different initial conditions. An application is presented to illustrate and interpret the estimator results.

## I. INTRODUCTION

THE KALMAN FILTER is the classical solution to the problem of estimating the state of a linear dynamic system in the presence of noise (see, e.g., [1], [2]). In the context of an information system, however, the estimator may be only one component of a complex structure that processes information from multiple sources to achieve end-user goals. For example, estimation is an essential component of many surveillance and situation assessment applications, and estimator outputs based upon sensor inputs may serve as inputs to data fusion and decision-making functions that often involve human judgment (perhaps synthetically, via expert systems) to formulate high-level evaluations. A concern is that, though the estimation approach may be considered “optimal” from an estimation-theoretic context, when considered as a component in a complex structure, the significance of the optimality claims may be reduced.

Many researchers have questioned the appropriateness of Bayesian-based decision theory as an inference mechanism for combining evidence or data to reach conclusions. Unfortunately, however, there seems to be no panacean alternative to probability theory for combining data for dynamic system state estimation, or filtering. The methods of stochastic systems are based largely upon Bayesian conditioning, and there do not seem to be counterparts in, say, Dempster-Shafer evidential reasoning or fuzzy logic. Thus, it may be difficult or impossible to abandon Bayesian methods for dynamic state estimation.

There are problems, however, with Bayesian-based filtering. One issue is the specification of initial conditions, particularly for collections that are data-limited. It is often the case that the *a priori* distribution used to initialize the

Kalman filter is not well known. If insufficient data are available to dampen the effect of incorrect initial conditions, biased estimates may result. Another issue is observability, such as may occur when the available data do not allow a unique determination of the system state, due, for example, to weak or fading signals. In such cases, the system may be only marginally or partially observable, and a classical (point-valued) estimator (e.g., the Kalman filter) will not provide an acceptable estimate, since it calculates only one of many possible state estimates that is consistent with the data. In this paper we provide a means of addressing these two key concerns within the context provided by the Bayesian approach.

## II. EPISTEMOLOGY

Convex Bayes decision theory [3] is an extension of classical probability-based decision theory. The philosophical basis underlying the development of convex Bayes decision theory is the epistemology of Levi [4], [5]. In this epistemology, the desideratum of an inquiry is to *avoid error*. The goal of classical estimation theory, however, is to minimize some error metric, such as the mean-squared error of the estimate. While we do not wish to abandon this classical goal, it is well known that classical estimation results, though perhaps optimal, are not robust with respect to changes in the modeling assumptions. Furthermore, the implementation of this criterion leads to a single estimate that is represented as being the “best” one, hence (it is assumed, at least implicitly) the only one worthy of serious consideration. The purpose of this development is to introduce a method for viewing classical estimation results in a larger context, resulting in an extension to the methodology.

The desideratum of avoiding error leads to a shift from the goal of determining a unique optimal estimate to a goal of finding a family of estimates, each of which has a claim to optimality under the appropriate assumptions. Such a shift is warranted in situations where there is insufficient knowledge to specify uniquely all of the parameters of the estimation rule.

Ignorance, in its root meaning, means lack of knowledge; uncertainty, on the other hand, typically means lack of precision. By specifying a probability distribution for a random variable, we attempt to characterize uncertainty. If the correct distribution function is unknown, that is a manifestation of ignorance. *Credal* probability is probability that reflects beliefs concerning the values of the state of the system, i.e., it characterizes precision. A *credal state* of a system is a family or set of credal probabilities, each of which is regarded as a seriously possible probabilistic description of the state of the system. The inability to choose between, say, two candidate distributions to characterize a random variable

Manuscript received April 6, 1990; revised July 28, 1990. This work was supported by ESL Inc., Sunnyvale, CA.

D. R. Morrell is with the Department of Electrical and Computer Engineering, Arizona State University, Tempe, AZ 85287-5706.

W. C. Stirling is with the Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT 84602.

IEEE Log Number 9039999.

0018-9472/91/0100-0184\$01.00 ©1991 IEEE

constitutes a form of conflict. One way to resolve this conflict is to permit all compromise solutions, in the form of all weighted averages of the two rival distributions, to be considered. Thus, it seems reasonable to suppose that a credal state should be a *convex set* of distributions. We may, therefore, consider an initial credal state of a system as being a convex set of *a priori* probabilities. The likelihood probabilities governing the statistical behavior of the observations, given the state, are credal probabilities concerning the interrelationships between the observations and the state values, and the posterior probabilities of the state given the observations are credal probabilities reflecting belief in the values of the state after the prior and conditional probabilities have been combined via Bayes rule.

If a credal state consists of exactly one probability, then a condition of *credal uniqueness* exists. Credal uniqueness is implicit with classical estimation theory, with the result that there is only one prior probability, only one likelihood probability, and only one posterior probability. Typically, a sequence of data is processed to obtain a sequence of posterior distributions, and the estimate is the conditional expectation given the observations. Convex estimation, on the other hand, provides a set of estimates, with each member of the set being associated with a particular distribution in the credal state. Thus, each estimate is consistent with the data, and has a legitimate claim to optimality. Convex estimation, therefore, represents an extension to classical point-valued estimation, leading to the concept of set-valued filtering and smoothing [6–8], which is developed in detail in this paper.

Set-valued estimation is related to the older concept of set-membership estimation. With set-membership estimation, however, a stochastic model of the system unknowns is not used. Rather, it is assumed that the unknowns belong to bounded sets; the estimator computes a set of all system states that are consistent with the observations and the sets bounding the system unknowns. The set-membership approach was first taken by Witsenhausen in the context of a general control problem [9], [10]. An approximate solution to the filtering problem, in which sets of possible system states are approximated by ellipsoids, was given by Schweppe for a linear system model [11]. Bertsekas and Rhodes solved the filtering problem exactly for a linear system and sets defined by an energy constraint, and also gave an ellipsoidal approximation algorithm that has some advantages over Schweppe's [12]. There is no discussion by Bertsekas and Rhodes about system observability, although it seems that their algorithm should perform correctly when the system is partially observable. Krener showed that with certain assumptions, the standard Kalman-Bucy filter can be interpreted as a set-membership filter [13].

Some previous work has been done in which a stochastic system model is combined with a set-membership specification of system unknowns (e.g., the initial state estimate is constrained to lie in a given set). Kats and Kurzanskii approached the problem of linear system estimation with arbitrary convex sets [14]. Pshenichnyi and Pokotilo have extended the formulation of Kats and Kurzanskii to allow the analysis of arbitrary sets with nonlinear systems, although they have apparently not provided a solution [15]. Anan'ev and Kurzanskii [16] present independent work that parallels this present study, and provide a solution for the best min-max conditional mean for the state of a nonlinear multistage system.

### III. SET-VALUED FILTERING

The notational convention for this development will be to use bold-face, lower-case symbols to denote random vectors, and sets of random vectors will be denoted by the corresponding bold-face, upper-case symbol. Sample-values of random vectors will be denoted by under-barred, italic symbols. Sets of sample-values will be denoted by the corresponding upper-case under-barred italic symbol. Also, matrices will be denoted by upper-case (non-under-barred) italic symbols.

Consider a discrete-time finite-dimensional system with state random vector  $x_t$ ,  $t = 0, 1, 2, \dots$ , representing the system state, and a sequence of random vectors  $z_t$  representing the observations of this system. We restrict attention to systems that can be modeled as first-order Markov processes, and assume that the observations errors are memoryless. We shall assume, in this development, that all distributions are absolutely continuous, hence probability densities exist. A probabilistic model of the system state may be expressed in terms of the following probability density functions.<sup>1</sup>

- The *transition* probability densities  $\{p_{x_t|x_{t-1}}(\cdot|\cdot), t = 0, 1, \dots\}$  that characterize the probability of state transitions from time  $t-1$  to time  $t$ .
- The *likelihood* probability densities  $\{p_{z_t|x_t}(\cdot|\cdot), t = 1, 2, \dots\}$  that characterize the probability governing the observation random vector  $z_t$  given state  $x_t$ .
- The *a priori*, or prior, probability density  $p_{x_0}(\cdot)$  that characterizes the probability distribution of the initial conditions of the system.
- The *a posteriori*, or posterior, probability density  $p_{x_t|Z_t}(\cdot|\cdot)$  that characterizes the probability distribution of the system state at time  $t$  given observations  $Z_t = \{z_1, \dots, z_t\}$ .

Classical filtering theory assumes credal uniqueness for all of these credal states. In this development, we retain the assumption of credal uniqueness for the transition and likelihood credal states, but reject it for the prior and, hence, the posterior, credal state.

Let  $t = 1, 2, \dots$ , and define  $Z_t = \{z_1, \dots, z_t\}$  as the finite sequence of observables up to time  $t$ . Suppose  $\underline{Z}_t = \{z_1, \dots, z_t\}$  are corresponding sample values of  $Z_t$ . We may invoke the Markov and memoryless properties of the system and apply Bayes rule to obtain the *posteriori* probability density  $p_{x_t|Z_t}(\xi|\underline{Z}_t)$ . Note that  $Z_t = \{Z_{t-1}, z_t\}$ . Let  $\mathcal{Y}_n$  denote the set of probability density functions defined over  $\mathcal{R}^n$ , and define the operators  $\Gamma_t^-: \mathcal{Y}_n \mapsto \mathcal{Y}_n$  and  $\Gamma_t^+: \mathcal{Y}_n \mapsto \mathcal{Y}_n$  as

$$\begin{aligned} & (\Gamma_t^- p_{x_{t-1}|Z_{t-1}})(\xi|\underline{Z}_{t-1}) \\ &= \int_{\mathcal{R}^n} p_{x_t|x_{t-1}}(\xi|\eta) p_{x_{t-1}|Z_{t-1}}(\eta|\underline{Z}_{t-1}) d\eta, \quad (1) \end{aligned}$$

and

$$\begin{aligned} & (\Gamma_t^+ p_{x_t|Z_{t-1}})(\xi|\underline{Z}_{t-1}, z_t) \\ &= \frac{p_{z_t|x_t}(z_t|\xi) p_{x_t|Z_{t-1}}(\xi|\underline{Z}_{t-1})}{\int_{\mathcal{R}^n} p_{z_t|x_t}(z_t|\eta) p_{x_t|Z_{t-1}}(\eta|\underline{Z}_{t-1}) d\eta}. \quad (2) \end{aligned}$$

<sup>1</sup>We adopt the standard notational practice of denoting the random variables associated with the density functions via subscripts.

Note that  $p_{x_t|Z_{t-1}} = \Gamma_t^- p_{x_{t-1}|Z_{t-1}}$ , and represents a time-update operation. Also,  $p_{x_t|Z_t} = \Gamma_t^+ p_{x_t|Z_{t-1}}$ , and represents a measurement-update operation. Define the operator  $\Gamma_{t|t-1}: \mathcal{X}_n \rightarrow \mathcal{X}_n$  by  $\Gamma_{t|t-1} = \Gamma_t^+ \circ \Gamma_t^-$ , i.e.,

$$(\Gamma_{t|t-1} p_{x_{t-1}|Z_{t-1}})(\xi|Z_t) = (\Gamma_t^+ (\Gamma_t^- p_{x_{t-1}|Z_{t-1}}))(\xi, Z_t), \quad (3)$$

and note that [17]

$$\begin{aligned} p_{x_t|Z_t}(\xi|Z_t) &= (\Gamma_{t|t-1} p_{x_{t-1}|Z_{t-1}})(\xi|Z_{t-1}, Z_t) \\ &= \frac{p_{z_t|x_t}(z_t|\xi) \int_{\mathcal{X}^n} p_{x_t|x_{t-1}}(\xi|\beta) p_{x_{t-1}|Z_{t-1}}(\beta|Z_{t-1}) d\beta}{\int_{\mathcal{X}^n} p_{z_t|x_t}(z_t|\gamma) \int_{\mathcal{X}^n} p_{x_t|x_{t-1}}(\gamma|\beta) p_{x_{t-1}|Z_{t-1}}(\beta|Z_{t-1}) d\beta d\gamma} \end{aligned} \quad (4)$$

for  $t=1, 2, \dots$ , represents a recursive time-and measurement updating equation for propagating the conditional probability densities  $p_{x_t|Z_t}$ . This recursion may be initialized by defining  $p_{x_0|Z_0}(\xi|Z_0) = p_{x_0}(\xi)$ .

#### A. The Credal States of a System

We shall say a set  $\mathcal{E} \subset \mathcal{X}_n$  is *convex* if, for any set of positive real numbers  $\{\alpha_i\}_{i=1}^k$ ,  $k \geq 1$ , such that  $\sum_{i=1}^k \alpha_i = 1$ , and any set of probability densities  $\{p_i\}_{i=1}^k \subset \mathcal{E}$ , then  $p^* = \sum_{i=1}^k \alpha_i p_i \in \mathcal{E}$ . A probability density of this form is said to be a *convex combination* of probability densities in  $\mathcal{E}$ . In particular, an *a priori* credal state is a convex set of prior probability densities of the form  $\mathcal{B}_{0|0} = \{p_{x_0}^\alpha, \alpha \in \mathcal{S}\}$ , where  $\mathcal{S}$  is an arbitrary index set.

The selection of a credal state is dependent upon the logical and contextual information that is available. For example, if there is sufficient information to choose a single distribution, then a credal state would consist of only one probability density. A strict Bayesian interpretation of the credal state is that it must contain one and only one prior probability density function—credal uniqueness. If we generalize to the notion of credal convexity and allow more than one prior to exist, then we must also permit more than one posterior probability density to exist. At time  $t$ , let  $\mathcal{B}_{t|t}$  denote a convex set of conditional probability densities of the form  $p_{x_t|Z_t}$ , for the system state at time  $t$ , given  $Z_t = Z_t$ .

For any set of probability densities  $\mathcal{Y}_{t-1|t-1} = \{p_{x_{t-1}|Z_{t-1}}\}$ ,  $t=1, 2, \dots$  ( $\mathcal{Y}_{t-1|t-1}$  may or may not be convex), we denote its image under the operator  $\Gamma_t^-$  by

$$\begin{aligned} \mathcal{Y}_{t|t-1} &= \Gamma_t^- \mathcal{Y}_{t-1|t-1} \\ &\stackrel{\text{def}}{=} \{p_{x_t|Z_{t-1}} = \Gamma_t^- p_{x_{t-1}|Z_{t-1}}; p_{x_{t-1}|Z_{t-1}} \in \mathcal{Y}_{t-1|t-1}\}, \end{aligned}$$

and we denote the image of  $\mathcal{Y}_{t|t-1}$  under the operator  $\Gamma_t^+$  by

$$\mathcal{Y}_{t|t} = \Gamma_t^+ \mathcal{Y}_{t|t-1} \stackrel{\text{def}}{=} \{p_{x_t|Z_t} = \Gamma_t^+ p_{x_t|Z_{t-1}}; p_{x_t|Z_{t-1}} \in \mathcal{Y}_{t|t-1}\}.$$

It follows that

$$\begin{aligned} \mathcal{Y}_{t|t} &= \Gamma_{t|t-1} \mathcal{Y}_{t-1|t-1} \\ &\stackrel{\text{def}}{=} \{p_{x_t|Z_t} = \Gamma_{t|t-1} p_{x_{t-1}|Z_{t-1}}; p_{x_{t-1}|Z_{t-1}} \in \mathcal{Y}_{t-1|t-1}\}. \end{aligned}$$

The convexity properties of the posterior distributions are summarized by the following *convex propagation theorem*.

**Theorem 1:** For a finite-dimensional discrete-state system characterized as a first-order Markov process with state transition probability densities  $\{p_{x_t|x_{t-1}}, t=0, 1, \dots\}$  and memoryless observations with likelihood probability densities  $\{p_{z_t|x_t}, t=0, 1, \dots\}$ , suppose the prior probability density  $p_{x_0}$  lies in a convex set  $\mathcal{B}_{0|0}$ . Then the set of posterior probability densities

$$\mathcal{B}_{t|t} = \{p_{x_t|Z_t} = \Gamma_{t|t-1}[p_{x_{t-1}|Z_{t-1}}]; p_{x_{t-1}|Z_{t-1}} \in \mathcal{B}_{t-1|t-1}\}$$

is convex: i.e., the operator  $\Gamma_{t|t-1}$  preserves convexity of the credal state.

To establish this theorem, we first must prove the following lemma.

**Lemma 1:** Let  $\{p_1, \dots, p_k\} \subset \mathcal{X}_n$ , and let  $\{\alpha_i\}_{i=1}^k$  be positive real numbers such that  $\sum_{i=1}^k \alpha_i = 1$ . Then there exist two sets of positive real numbers  $\{\alpha_i\}_{i=1}^k$  and  $\{b_i\}_{i=1}^k$ , such that  $\sum_{i=1}^k \alpha_i = \sum_{i=1}^k b_i = 1$  and

$$\Gamma_t^+ \left[ \sum_{i=1}^k \alpha_i p_i \right] = \sum_{i=1}^k \alpha_i \Gamma_t^+ p_i \quad (5)$$

$$\sum_{i=1}^k \alpha_i \Gamma_t^+ p_i = \Gamma_t^+ \left[ \sum_{i=1}^k b_i p_i \right]. \quad (6)$$

*Proof of Lemma 1:* We first show that (5) holds:

$$\begin{aligned} \left( \Gamma_t^+ \left[ \sum_{i=1}^k \alpha_i p_i \right] \right) (\xi|\xi) &= \frac{p_{z_t|x_t}(\xi|\xi) \left[ \sum_{i=1}^k \alpha_i p_i(\xi) \right]}{\int_{\mathcal{X}^n} p_{z_t|x_t}(\xi|\gamma) \left[ \sum_{j=1}^k \alpha_j p_j(\gamma) \right] d\gamma} \\ &= \frac{\sum_{i=1}^k \alpha_i p_{z_t|x_t}(\xi|\xi) p_i(\xi)}{\sum_{j=1}^k \alpha_j \int_{\mathcal{X}^n} p_{z_t|x_t}(\xi|\gamma) p_j(\gamma) d\gamma}. \end{aligned} \quad (7)$$

We define  $\pi_i = \int_{\mathcal{X}^n} p_{z_t|x_t}(\xi|\gamma) p_i(\gamma) d\gamma$  and assume without loss of generality that  $\pi_i > 0$ .<sup>2</sup> Equation (7) becomes

$$\begin{aligned} \left( \Gamma_t^+ \left[ \sum_{i=1}^k \alpha_i p_i \right] \right) (\xi|\xi) &= \sum_{i=1}^k \frac{\alpha_i}{\left[ \sum_{j=1}^k \alpha_j \pi_j \right]} p_{z_t|x_t}(\xi|\xi) p_i(\xi) \\ &= \sum_{i=1}^k \left[ \frac{\alpha_i \pi_i}{\sum_{j=1}^k \alpha_j \pi_j} \right] \left[ \frac{p_{z_t|x_t}(\xi|\xi) p_i(\xi)}{\pi_i} \right]. \end{aligned} \quad (8)$$

<sup>2</sup>If  $\pi_i = 0$ , then  $p_{z_t|x_t}(\xi|\gamma) p_i(\gamma) = 0$  for all  $\gamma$ , and the  $i$ th term may be removed from the summations in both the numerator and denominator of (7) without affecting their values.

Let  $a_i = \alpha_i \pi_i / \sum_{j=1}^k \alpha_j \pi_j$ , and note that  $\sum_{i=1}^k a_i = 1$  and  $a_i > 0$ ,  $i \in \{1, \dots, k\}$ . Using the expressions for  $\pi_i$  and  $a_i$  in (8),

$$\begin{aligned} \left( \Gamma_t^+ \left[ \sum_{i=1}^k \alpha_i p_i \right] \right) (\underline{\xi} | \underline{\zeta}) &= \sum_{i=1}^k a_i \frac{p_{z_i|x_i}(\underline{\zeta} | \underline{\xi}) p_i(\underline{\xi})}{\int_{\mathcal{X}^n} p_{z_i|x_i}(\underline{\zeta} | \underline{\gamma}) p_i(\underline{\gamma}) d\underline{\gamma}} \\ &= \sum_{i=1}^k a_i (\Gamma_t^+ p_i) (\underline{\xi}). \end{aligned} \quad (9)$$

We now prove (6):

$$\begin{aligned} \sum_{i=1}^k \alpha_i (\Gamma_t^+ p_i) (\underline{\xi} | \underline{\zeta}) &= \sum_{i=1}^k \alpha_i \left[ \frac{p_{z_i|x_i}(\underline{\zeta} | \underline{\xi}) p_i(\underline{\xi})}{\int_{\mathcal{X}^n} p_{z_i|x_i}(\underline{\zeta} | \underline{\gamma}) p_i(\underline{\gamma}) d\underline{\gamma}} \right] \\ &= \sum_{i=1}^k \alpha_i \left[ \frac{p_{z_i|x_i}(\underline{\zeta} | \underline{\xi}) p_i(\underline{\xi})}{\pi_i} \right]. \end{aligned} \quad (10)$$

Let  $b_i = (\alpha_i / \pi_i) / \sum_{j=1}^k (\alpha_j / \pi_j)$ . Note that  $\sum_{i=1}^k b_i = 1$  and  $b_i > 0$ ,  $i \in \{1, \dots, k\}$ . It may be verified directly that  $\alpha_i = b_i \pi_i / \sum_{j=1}^k b_j \pi_j$ . We may now rewrite (10) as

$$\sum_{i=1}^k \alpha_i (\Gamma_t^+ p_i) (\underline{\xi} | \underline{\zeta}) = \sum_{i=1}^k \frac{\alpha_i}{b_i \pi_i} p_{z_i|x_i}(\underline{\zeta} | \underline{\xi}) b_i p_i(\underline{\xi}). \quad (11)$$

Substituting for  $\alpha_i$  and  $\pi_i$  in (11) gives

$$\begin{aligned} \sum_{i=1}^k \alpha_i (\Gamma_t^+ p_i) (\underline{\xi} | \underline{\zeta}) &= \sum_{i=1}^k \frac{p_{z_i|x_i}(\underline{\zeta} | \underline{\xi}) b_i p_i(\underline{\xi})}{\sum_{j=1}^k b_j \int_{\mathcal{X}^n} p_{z_j|x_j}(\underline{\zeta} | \underline{\gamma}) p_j(\underline{\gamma}) d\underline{\gamma}} \\ &= \frac{p_{z_i|x_i}(\underline{\zeta} | \underline{\xi}) \left[ \sum_{i=1}^k b_i p_i(\underline{\xi}) \right]}{\int_{\mathcal{X}^n} p_{z_i|x_i}(\underline{\zeta} | \underline{\gamma}) \left[ \sum_{j=1}^k b_j p_j(\underline{\gamma}) \right] d\underline{\gamma}} \\ &= \left( \Gamma_t^+ \left[ \sum_{i=1}^k b_i p_i \right] \right) (\underline{\xi} | \underline{\zeta}). \end{aligned} \quad (12)$$

*Proof of Theorem 1:* The proof is by induction. By hypothesis,  $\mathcal{B}_{0|0}$  is convex. We assume that  $\mathcal{B}_{t-1|t-1}$  is convex, and must show that  $\mathcal{B}_{t|t}$  is convex. Note that  $\mathcal{B}_{t|t-1} = \Gamma_t^- \mathcal{B}_{t-1|t-1}$  and  $\mathcal{B}_{t|t} = \Gamma_t^+ \mathcal{B}_{t|t-1}$ . It is obvious that  $\Gamma_t^-$  preserves convexity since it is a linear operator. Thus,  $\mathcal{B}_{t|t-1}$  is convex, and it remains to show that  $\Gamma_t^+$  preserves convexity. Consider two probability densities  $p_1$  and  $p_2$  arbitrarily chosen from  $\mathcal{B}_{t|t}$ . Since  $\mathcal{B}_{t|t}$  is the image of  $\mathcal{B}_{t|t-1}$  under  $\Gamma_t^+$ , there are two probability densities  $p_1^-, p_2^- \in \mathcal{B}_{t|t-1}$  such that  $p_1 = \Gamma_t^+ p_1^-$  and  $p_2 = \Gamma_t^+ p_2^-$ . Let  $p_*$  be defined as  $p_* = \alpha p_1 + (1-\alpha) p_2$ , where  $\alpha \in (0, 1)$ . It must be shown that  $p_* \in \mathcal{B}_{t|t}$ . From Lemma 1, there is an  $\alpha \in [0, 1]$  such that

$$p_* (\underline{\xi} | \underline{\zeta}) = (\Gamma_t^+ [\alpha p_1^- + (1-\alpha) p_2^-]) (\underline{\xi} | \underline{\zeta}). \quad (13)$$

Thus, by convexity of  $\mathcal{B}_{t|t-1}$ ,  $\alpha p_1^- + (1-\alpha) p_2^- \in \mathcal{B}_{t|t-1}$  and, since  $\mathcal{B}_{t|t}$  is the image of  $\mathcal{B}_{t|t-1}$  under  $\Gamma_t^+$ ,  $p_* \in \mathcal{B}_{t|t}$ . Thus, since  $\Gamma_{t|t-1} = \Gamma_t^+ \circ \Gamma_t^-$ , convexity is established.

## B. Linear Gaussian Systems

Consider a linear stochastic system of the form

$$\mathbf{x}_t = F_{t-1} \mathbf{x}_{t-1} + G_{t-1} \mathbf{w}_{t-1} \quad (14)$$

$$\mathbf{z}_t = H_t \mathbf{x}_t + \mathbf{v}_t, \quad (15)$$

for  $t = 1, 2, \dots$ , where  $F_t$  is  $n \times n$ ,  $G_t$  is  $n \times p$ ,  $H_t$  is  $m \times n$ ,  $\{\mathbf{w}_t\}$  and  $\{\mathbf{v}_t\}$  are  $p$ - and  $m$ -dimensional vector Gaussian (zero-mean) white noise processes with positive-definite covariance matrices  $Q_t$  and  $R_t$ , respectively.

The conventional modeling approach is to assume credal uniqueness, and characterize the initial state as being distributed  $\mathbf{x}_0 \sim \mathcal{N}(\hat{\mathbf{x}}_{0|0}, P_{0|0})$  (i.e.,  $\mathbf{x}_0$  is an  $n$ -dimensional Gaussian random vector with mean  $\hat{\mathbf{x}}_{0|0}$  and covariance  $P_{0|0}$ .) We wish to relax this assumption, and specify an initial credal state,  $\mathcal{B}_{0|0}$ , consisting of a convex set of prior distributions. Suppose we have information that the prior distribution is Gaussian with known covariance  $P_{0|0}$  and with mean value lying in a convex set in state space defined by

$$\underline{X}_{0|0} = \{ \mathbf{x} \in \mathcal{R}^n : \| K_{0|0}^{-1} (\mathbf{x} - \mathbf{c}_{0|0}) \| \leq 1 \},$$

where  $K_{0|0}$  is a nonsingular matrix. If we use the Euclidean norm, then

$$\underline{X}_{0|0} = \{ \mathbf{x} \in \mathcal{R}^n : [\mathbf{x} - \mathbf{c}_{0|0}]^T S_{0|0}^{-2} [\mathbf{x} - \mathbf{c}_{0|0}] \leq 1 \}, \quad (16)$$

where  $S_{0|0} = K_{0|0} K_{0|0}^T$ . (Note that if  $K_{0|0} = 0$ , then  $\underline{X}_{0|0} = \{\mathbf{c}_{0|0}\}$  is a singleton set, and the representation degenerates to the conventional modeling approach.) With this norm, the set  $\underline{X}_{0|0}$  is a hyper-ellipsoidal region in  $\mathcal{R}^n$  centered at  $\mathbf{c}_{0|0}$  with semi-axes defined by the eigenvalues and eigenvectors of  $S_{0|0}^{-1}$ , and contains the mean values of the set of random variables  $\mathcal{X}_{0|0} = \{ \mathbf{x} \sim \mathcal{N}(\mathbf{x}, P_{0|0}) : \mathbf{x} \in \underline{X}_{0|0} \}$ . Let  $\mathcal{V}_{0|0} = \{ \mathcal{N}(\mathbf{x}, P_{0|0}) : \mathbf{x} \in \underline{X}_{0|0} \}$ , and note that  $\mathcal{V}_{0|0}$  is not a convex set, since convex combinations of Gaussian distributions are not Gaussian.

For any set of probability densities  $\mathcal{V}$ , we may form a convex set of probability densities  $\mathcal{B}$  by taking the convex closure of  $\mathcal{V}$ , denoted  $\overline{\mathcal{V}}$ , as the smallest set containing all finite convex combinations of elements of  $\mathcal{V}$ . The set  $\mathcal{V}$  is then said to be a *generator set* for  $\mathcal{B}$ . Thus, we form the convex closure  $\overline{\mathcal{V}}_{0|0}$  of  $\mathcal{V}_{0|0}$ . Then  $\mathcal{B}_{0|0} = \overline{\mathcal{V}}_{0|0}$  is the *a priori* credal state of the system. This procedure however, violates the assumption that  $\mathbf{x}_0$  is governed by a Gaussian distribution. We are strongly motivated to retain the Gaussian hypothesis since, if the transition, likelihood, and prior distributions are all Gaussian, then the posterior distribution given by (4) is also Gaussian. Still, the notion of convexity is an extremely useful way to characterize the state of knowledge about the system, and we are reluctant to discard it. Fortunately, we may reconceive this problem by the following convex representation theorem.

*Theorem 2:* Let  $\mathcal{V}_{t|t-1} \in \mathcal{V}_n$  be an arbitrary set of probability densities. Then the convex closure of  $\Gamma_{t|t-1} \mathcal{V}_{t|t-1}$  is equal to  $\Gamma_{t|t-1}$  operating on the convex closure of  $\mathcal{V}_{t-1|t-1}$ , i.e.,

$$\overline{\Gamma_{t|t-1} \mathcal{V}_{t|t-1}} = \Gamma_{t|t-1} \overline{\mathcal{V}_{t-1|t-1}}. \quad (17)$$

*Proof:* We note that

$$\mathcal{B}_{t|t-1} = \Gamma_t^- \mathcal{B}_{t-1|t-1}, \mathcal{B}_{t|t} = \Gamma_t^+ \mathcal{B}_{t|t-1}, \mathcal{Y}_{t|t-1} = \Gamma_t^- \mathcal{Y}_{t-1|t-1},$$

and  $\mathcal{Y}_{t|t} = \Gamma_t^+ \mathcal{Y}_{t|t-1}$ . We first prove that the desired relationship holds under  $\Gamma_t^-$ . Let  $\mathcal{B}_{t-1|t-1} = \mathcal{Y}_{t-1|t-1}$  and  $\mathcal{E} = \mathcal{Y}_{t|t-1}$ . To prove that  $\mathcal{B}_{t|t-1} \subseteq \mathcal{E}$ , let  $p_*^- \in \mathcal{B}_{t|t-1}$ . Then there exists a  $p_* \in \mathcal{B}_{t-1|t-1}$  such that  $p_*^- = \Gamma_t^- p_*$ . Since  $\mathcal{B}_{t-1|t-1}$  is the set of all finite convex combinations of the elements of  $\mathcal{Y}_{t-1|t-1}$ ,  $p_* = \sum_{i=1}^k \alpha_i p_i$ , where  $\alpha_i > 0$  and  $p_i \in \mathcal{Y}_{t-1|t-1}$ ,  $i \in \{1, \dots, k\}$ , and  $\sum_{i=1}^k \alpha_i = 1$ . Now, by the linearity of  $\Gamma_t^-$ ,

$$p_*^- = \Gamma_t^- p_* = \Gamma_t^- \left[ \sum_{i=1}^k \alpha_i p_i \right] = \sum_{i=1}^k \alpha_i (\Gamma_t^- p_i) = \sum_{i=1}^k \alpha_i p_i^-, \quad (18)$$

where  $p_i^- = \Gamma_t^- p_i \in \mathcal{Y}_{t|t-1}$ . Thus,  $p_*^- \in \mathcal{E}$  and  $\mathcal{B}_{t|t-1} \subseteq \mathcal{E}$ .

To prove that  $\mathcal{E} \subseteq \mathcal{B}_{t|t-1}$ , let  $p_*^- \in \mathcal{E}$ . Then  $p_*^- = \sum_{i=1}^k \alpha_i p_i^-$ , where  $\alpha_i > 0$  and  $p_i^- \in \mathcal{Y}_{t|t-1}$ ,  $i \in \{1, \dots, k\}$ , and  $\sum_{i=1}^k \alpha_i = 1$ . Since  $\mathcal{Y}_{t|t-1}$  is the image of  $\mathcal{Y}_{t-1|t-1}$  under  $\Gamma_t^-$ , for each  $p_i^- \in \mathcal{Y}_{t|t-1}$  there is a  $p_i \in \mathcal{Y}_{t-1|t-1}$  such that  $p_i^- = \Gamma_t^- p_i$ . Thus by the linearity of  $\Gamma_t^-$ ,

$$p_*^- = \sum_{i=1}^k \alpha_i (\Gamma_t^- p_i) = \Gamma_t^- \left[ \sum_{i=1}^k \alpha_i p_i \right] = \Gamma_t^- p_*, \quad (19)$$

where  $p_* = \sum_{i=1}^k \alpha_i p_i \in \mathcal{B}_{t-1|t-1}$ . Then, since  $p_*^- \in \mathcal{B}_{t|t-1}$ ,  $\mathcal{E} \subseteq \mathcal{B}_{t|t-1}$ .

We next prove that the relationship holds under the  $\Gamma_t^+$  operation. Let  $\mathcal{E}$  be the set of all finite convex combinations of the elements of  $\mathcal{Y}_{t|t}$ . To prove that  $\mathcal{B}_{t|t} \subseteq \mathcal{E}$ , let  $p_*^+ \in \mathcal{B}_{t|t}$ . Then there exists a  $p_*^- \in \mathcal{B}_{t|t-1}$  such that  $p_*^+ = \Gamma_t^+ p_*^-$ . Since  $\mathcal{B}_{t|t-1}$  is the set of all finite convex combinations of the elements of  $\mathcal{Y}_{t|t-1}$ ,  $p_*^- = \sum_{i=1}^k \alpha_i p_i^-$ , where  $\alpha_i > 0$  and  $p_i^- \in \mathcal{Y}_{t|t-1}$ ,  $i \in \{1, \dots, k\}$ , and  $\sum_{i=1}^k \alpha_i = 1$ . From Lemma 1, there is a set of positive real numbers  $\{a_i\}$ ,  $i \in \{1, \dots, k\}$  such that  $\sum_{i=1}^k a_i = 1$  and

$$p_*^+ = \Gamma_t^+ p_*^- = \Gamma_t^+ \left[ \sum_{i=1}^k \alpha_i p_i^- \right] = \sum_{i=1}^k a_i (\Gamma_t^+ p_i^-) = \sum_{i=1}^k a_i p_i^+, \quad (20)$$

where  $p_i^+ = \Gamma_t^+ p_i^- \in \mathcal{Y}_{t|t}$ . Thus,  $p_*^+ \in \mathcal{E}$  and  $\mathcal{B}_{t|t} \subseteq \mathcal{E}$ .

To prove that  $\mathcal{E} \subseteq \mathcal{B}_{t|t}$ , let  $p_*^+ \in \mathcal{E}$ . Then  $p_*^+ = \sum_{i=1}^k b_i p_i^+$ , where  $b_i > 0$  and  $p_i^+ \in \mathcal{Y}_{t|t}$ ,  $i \in \{1, \dots, k\}$ , and  $\sum_{i=1}^k b_i = 1$ . Since  $\mathcal{Y}_{t|t}$  is the image of  $\mathcal{Y}_{t|t-1}$  under  $\Gamma_t^+$ , for each  $p_i^+ \in \mathcal{Y}_{t|t}$  there is a  $p_i^- \in \mathcal{Y}_{t|t-1}$  such that  $p_i^+ = \Gamma_t^+ p_i^-$ . From Lemma 1, there exists a set  $\{\beta_i\}$ ,  $i \in \{1, \dots, k\}$ , such that  $\sum_{i=1}^k \beta_i = 1$  and

$$p_*^+ = \sum_{i=1}^k b_i (\Gamma_t^+ p_i^-) = \Gamma_t^+ \left[ \sum_{i=1}^k \beta_i p_i^- \right]. \quad (21)$$

Since  $\mathcal{B}_{t|t-1}$  is the set of finite convex combinations of  $\mathcal{Y}_{t|t-1}$ ,  $\sum_{i=1}^k \beta_i p_i^- \in \mathcal{B}_{t|t-1}$ . Then since  $\mathcal{B}_{t|t}$  is the image of  $\mathcal{B}_{t|t-1}$  under  $\Gamma_t^+$ ,  $p_*^+ \in \mathcal{B}_{t|t}$ , and  $\mathcal{E} \subseteq \mathcal{B}_{t|t}$ . This establishes the result.

We may obtain  $\mathcal{B}_{t|t}$  by performing the Bayesian update on elements of  $\mathcal{Y}_{t|t}$  and then taking the convex combination. There are a number of practical reasons for restricting consideration only to a set of Gaussian probability densities.

- Gaussian models are appropriate for many physical systems, due to considerations such as the central limit theorem. It is often wise (for example, based upon

considerations of maximum entropy), to assume the Gaussian model unless there is explicit evidence to the contrary. Also, linear systems driven by Gaussian inputs yield Gaussian outputs, so if the prior is Gaussian and all noise sources are Gaussian, then the output will be Gaussian. Although neither of these reasons represent blanket warrants for assuming the Gaussian hypothesis, Gaussian models have withstood the test of application successfully with great regularity.

- The restriction of the prior credal set to Gaussian probability densities with common covariance but whose means lie in a convex set makes it possible to replace the notion of a convex set of probability densities with a convex set of means. Since conventional estimation techniques (such as the Kalman filter) admit an interpretation as the conditional expectation of the posterior distribution of the state given the observations, we may consider the development of a set-valued Kalman filter that computes the conditional expectation of a set  $\mathcal{Y}_{t|t}$  of posterior distributions of the state given the observations.

Even though convexity may allow us to consider all probability densities in  $\mathcal{B}_{t|t}$ , we may find it adequate to restrict attention only to the subset of Gaussian probability densities,  $\mathcal{Y}_{t|t}$ . Note that each element of  $\mathcal{Y}_{t|t}$  is Gaussian and, therefore, for any  $p_{x_t|z_t} \in \mathcal{Y}_{t|t}$ , we may apply the Kalman filter and obtain the predicted and filtered conditional probability densities via (4).

Given

$$p_{x_{t-1}|z_{t-1}} = \mathcal{N}(\hat{x}_{t-1|t-1}, P_{t-1|t-1}),$$

the predicted probability density is

$$\Gamma_t^- p_{x_t|z_{t-1}} = \mathcal{N}(\hat{x}_{t|t-1}, P_{t|t-1}),$$

where

$$\hat{x}_{t|t-1} = F_{t-1} \hat{x}_{t-1|t-1} \quad (22)$$

$$P_{t|t-1} = F_{t-1} P_{t-1|t-1} F_{t-1}^T + G_{t-1} Q_{t-1} G_{t-1}^T, \quad (23)$$

and the filtered probability density is

$$\Gamma_t^+ p_{x_t|z_t} = \mathcal{N}(\hat{x}_{t|t}, P_{t|t}),$$

where

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + W_t [z_t - H_t \hat{x}_{t|t-1}] \quad (24)$$

$$P_{t|t} = [I - W_t H_t] P_{t|t-1}, \quad (25)$$

with  $W_t$  the Kalman gain defined by

$$W_t = P_{t|t-1} H_t^T [H_t P_{t|t-1} H_t^T + R_t]^{-1}. \quad (26)$$

From these equations it is easy to see that  $\hat{x}_{t-1|t-1}$  and  $\hat{x}_{t|t}$  are linearly related, i.e.,

$$\hat{x}_{t|t} = A \hat{x}_{t-1|t-1} + \underline{b}, \quad (27)$$

where

$$A = [I - W_t H_t] F_{t-1} \quad (28)$$

$$\underline{b} = W_t z_t. \quad (29)$$

## IV. THE SET-VALUED KALMAN FILTER

In this development, we will assume credal convexity, and assume that the set of *a priori* distributions is the convex closure of  $\mathcal{Y}_{0|0} = \{\mathcal{N}(\underline{x}, P_{0|0}) : \underline{x} \in \underline{X}_{0|0}\}$ , where  $\underline{X}_{0|0}$  is defined by (16). Theorem 2 relieves us of the necessity, however, of dealing with all elements of this convex set. Instead, we may solve the estimation problem by deriving equations of evolution for elements only of  $\mathcal{Y}_{t|t}$ , and taking the convex closure to obtain  $\mathcal{B}_{t|t}$ . Consequently, we shall limit our discussion to the set  $\mathcal{Y}_{t|t}$  of Gaussian probability densities. Even with this simplification, if the set of prior distributions  $\mathcal{Y}_{0|0}$  is not a singleton set (i.e.,  $K_{0|0} \neq 0$ ), then it will be necessary, theoretically, to formulate a Kalman filter of the form given by (22)–(26) for each *a priori* distribution in  $\mathcal{Y}_{0|0}$ . The set-valued Kalman filter provides a means of doing this without implementing a continuum of Kalman filters.

Conventional linear Gaussian point-valued estimation takes the mean of the conditional distribution  $p_{x_t|z_t}$  as the filtered estimate of the system given by (14) and (15). This conditional probability density is obtained via (4) for each  $t$  under the assumption of credal uniqueness. With the relaxation of credal uniqueness, we assume that the initial state  $x_0$  is governed by a probability density from the set  $\mathcal{Y}_{0|0} = \{\mathcal{N}(\underline{x}, P_{0|0}) : \underline{x} \in \underline{X}_{0|0}\}$ , each of which is a seriously possible probabilistic description of  $x_0$ . For  $t = 1, 2, \dots$ , we wish to determine a set of probability densities of the form

$$\mathcal{Y}_{t|t} = \{\mathcal{N}(\underline{x}, P_{t|t}) : \underline{x} \in \underline{X}_{t|t}\}; \quad (30)$$

we show subsequently in Lemma 2 that  $\underline{X}_{t|t}$  has the form

$$\underline{X}_{t|t} \equiv \{\underline{x} : \|K_{t|t}^{-1}(\underline{x} - \underline{c}_{t|t})\| \leq 1\}. \quad (31)$$

The norm  $\|\cdot\|$  determines the basic shape of  $\underline{X}_{t|t}$  (i.e., ellipsoidal, rectangular, etc.),  $K_{t|t}$  determines the size and extent of  $\underline{X}_{t|t}$ , and  $\underline{c}_{t|t}$  is the *centroid* of  $\underline{X}_{t|t}$ . If the Euclidean norm is used (it need not be), (31) becomes

$$\underline{X}_{t|t} = \{\underline{x} \in \mathcal{R}^n : [\underline{x} - \underline{c}_{t|t}]^T S_{t|t}^{-1} [\underline{x} - \underline{c}_{t|t}] \leq 1\}, \quad (32)$$

where  $S_{t|t} = K_{t|t} K_{t|t}^T$ . The set-valued Kalman filtering problem, therefore, is to determine expressions for  $K_{t|t}$  and  $\underline{c}_{t|t}$ . Once these quantities are obtained, the set  $\underline{X}_{t|t}$  represents the set of mean-values of the conditional distributions in  $\mathcal{Y}_{t|t}$ , the set of all Gaussian probability densities derived from (4). Let  $\mathcal{A}$  denote the collection of all finite sets of numbers  $\{\alpha_i\}_{i=1}^k$  such that  $\alpha_i > 0$  and  $\sum_{i=1}^k \alpha_i = 1$ , for all integers  $k \geq 1$ . The credal state is

$$\mathcal{B}_{t|t} = \left\{ p_{x_t|z_t} = \sum_{i=1}^k \alpha_i \mathcal{N}(x_t, P_{t|t}) : x_t \in \underline{X}_{t|t}, \{\alpha_i\}_{i=1}^k \subset \mathcal{A} \right\}, \quad (33)$$

Thus, the credal state is the set of all finite Gaussian sums of elements of  $\mathcal{Y}_{t|t}$  and is, therefore, a very large collection of probability densities. Let  $p_1 \in \mathcal{Y}_{t|t}$  and  $p_2 \in \mathcal{Y}_{t|t}$ . Since mathematical expectation is a linear operation, the expected value of the convex combination  $p^* = \alpha_1 p_1 + \alpha_2 p_2$  is  $\hat{x}^* = \alpha_1 \hat{x}_1 + \alpha_2 \hat{x}_2$ , where  $\hat{x}_1$  and  $\hat{x}_2$  are the expected values of  $p_1$  and  $p_2$ , respectively. Consequently, the set  $\underline{X}_{t|t}$  consists of the expectations of all  $p_{x_t|z_t} \in \mathcal{B}_{t|t}$ . We shall denote  $\underline{X}_{t|t}$  as the *filtered credal set*.

For each  $\hat{x}_{t-1|t-1} \in \underline{X}_{t-1|t-1}$  there corresponds an  $\hat{x}_{t|t} \in \underline{X}_{t|t}$  through the linear relation defined by (27). In particular,

the centroid of  $\underline{X}_{t-1|t-1}$  maps to the centroid of  $\underline{X}_{t|t}$ , i.e.,

$$\underline{c}_{t|t} = F_{t-1} \underline{c}_{t-1|t-1} + W_t [\underline{z}_t - H_t F_{t-1} \underline{c}_{t-1|t-1}]. \quad (34)$$

It remains to determine an expression for  $K_{t|t}$  in terms of  $K_{t-1|t-1}$ . The following lemma establishes this relationship.

*Lemma 2:* Let  $\underline{X} = \{\underline{x} \in \mathcal{R}^n : \|K^{-1}(\underline{x} - \underline{c})\| \leq 1\}$  for some non-singular  $n$ -dimensional matrix  $K$  and an arbitrary  $n$ -vector  $\underline{c}$ . Then the image of  $\underline{X}$  under the mapping  $T(\underline{x}) = A\underline{x} + \underline{b}$  where  $A$  is a non-singular  $n$ -dimensional matrix and  $\underline{b}$  is an arbitrary  $n$ -vector is

$$\underline{Y} = \{\underline{y} \in \mathcal{R}^n : \|K^{-1}A^{-1}[\underline{y} - T(\underline{c})]\| \leq 1\}.$$

*Proof of Lemma 2:* Let  $\underline{x} \in \underline{X}$ . Then, by direct substitution,

$$\|K^{-1}A^{-1}[T(\underline{x}) - T(\underline{c})]\| = \|K^{-1}(\underline{x} - \underline{c})\| \leq 1,$$

so  $T(\underline{x}) \in \underline{Y}$ . Now, suppose  $\underline{y} \in \underline{Y}$ . We must show there exists an  $\underline{x} \in \underline{X}$  such that  $\underline{y} = T(\underline{x})$ . Let  $\underline{x} = A^{-1}(\underline{y} - \underline{b})$ . Then

$$\|K^{-1}(\underline{x} - \underline{c})\| = \|K^{-1}[A^{-1}(\underline{y} - \underline{b}) - \underline{c}]\|$$

but  $\|K^{-1}[A^{-1}(\underline{y} - \underline{b}) - \underline{c}]\| = \|K^{-1}[\underline{x} - T(\underline{c})]\| \leq 1$ , since  $\underline{y} \in \underline{Y}$ . This proves that  $\underline{Y}$  is the image of  $\underline{X}$  under  $T(\cdot)$ .

We apply Lemma 2 by setting  $\underline{X} = \underline{X}_{t-1|t-1}$  and letting  $A$  and  $\underline{b}$  be given by (28) and (29), respectively. Then  $\underline{X}_{t|t} = \underline{Y}$  is the image of  $\underline{X}_{t-1|t-1}$  under the resulting mapping, and

$$\begin{aligned} K_{t|t} &= AK_{t-1|t-1} \\ &= [I - W_t H_t] F_{t-1} K_{t-1|t-1}. \end{aligned} \quad (35)$$

We shall term the matrix  $K_{t|t}$  the *filtered credal matrix*.

We summarize the derivation of the set-valued Kalman filter with the following theorem:

*Theorem 3:* Consider the system

$$\begin{aligned} \mathbf{x}_t &= F_{t-1} \mathbf{x}_{t-1} + G_{t-1} \mathbf{w}_{t-1} \\ \mathbf{z}_t &= H_t \mathbf{x}_t + \mathbf{v}_t, \end{aligned}$$

$t = 1, 2, \dots$ , with  $\{\mathbf{w}_t\}$  and  $\{\mathbf{v}_t\}$  Gaussian (zero-mean) white noise processes with covariance  $Q_t$  and  $R_t$ , respectively. Let  $K_{0|0}$  be an invertible  $n \times n$  matrix, and let  $\underline{c}_{0|0}$  be an  $n$ -vector. Let  $\{\underline{z}_1, \underline{z}_2, \dots\}$  be samples of  $\{z_1, z_2, \dots\}$ . Suppose the initial state vector  $x_0$  is a member of the set of random vectors

$$\underline{X}_{0|0} = \{\mathbf{x} \sim \mathcal{N}(\underline{x}, P_{0|0}) : \underline{x} \in \underline{X}_{0|0}\}$$

where

$$\underline{X}_{0|0} = \{\underline{x} \in \mathcal{R}^n : \|K_{0|0}^{-1}(\underline{x} - \underline{c}_{0|0})\| \leq 1\},$$

and  $P_{0|0}$  is a positive-definite matrix.

The set-valued Kalman filter is as follows.

*Prediction Step:*

$$\underline{X}_{t|t-1} = \{\underline{x} \in \mathcal{R}^n : \|K_{t|t-1}^{-1}(\underline{x} - \underline{c}_{t|t-1})\| \leq 1\},$$

where

$$\begin{aligned} \underline{c}_{t|t-1} &= F_{t-1} \underline{c}_{t-1|t-1} \\ P_{t|t-1} &= F_{t-1} P_{t-1|t-1} F_{t-1}^T + G_{t-1} Q_{t-1} G_{t-1}^T \\ K_{t|t-1} &= F_{t-1} K_{t-1|t-1}. \end{aligned}$$

Filter Step:

$$\underline{X}_{t|t} = \{x \in \mathcal{R}^n: \|K_{t|t}^{-1}(x - c_{t|t})\| \leq 1\},$$

where

$$c_{t|t} = c_{t|t-1} + W_t [z_t - H_t c_{t|t-1}]$$

$$P_{t|t} = [I - W_t H_t] P_{t|t-1}$$

$$K_{t|t} = [I - W_t H_t] K_{t|t-1}$$

with  $W_t$  the Kalman gain defined by

$$W_t = P_{t|t-1} H_t^T [H_t P_{t|t-1} H_t^T + R_t]^{-1}.$$

The structure of the credal set is governed by the credal matrix. If  $K_{t|t} \rightarrow 0$  as  $t \rightarrow \infty$ , then the set estimates converges asymptotically to the point estimate  $c_{t|t}$ , and all prior distributions lead asymptotically to the same limiting posterior distribution. In such cases, the system is said to be *credible*. Examination of (35), (23), and (25) reveals that the credal matrix and the filtered covariance evolve according to the same dynamics, i.e., through  $[I - W_t H_t] F_{t-1}$ . It is well known that, for uniformly observable and controllable systems, this matrix is stable. Consequently, all uniformly observable and controllable systems are credible. Also, all detectable systems are credible.

#### V. SET-VALUED SMOOTHING

Development of the fixed-interval set-valued smoother follows the approach taken by Rauch-Tung-Streifel [18]. Let  $Z_T = \{z_1, \dots, z_T\}$  be a finite sample sequence. Each element of the smoothed credal set must propagate according to the equations

$$\hat{x}_{t|T} = C_t \hat{x}_{t+1|T} + (I - C_t F_t) \hat{x}_{t|t}, \quad (36)$$

$$P_{t|T} = P_{t|t} + C_t (P_{t+1|T} - P_{t+1|t}) C_t^T, \quad (37)$$

where  $C_t$ , the smoother correction matrix, is an  $n \times n$  matrix

$$C_t = P_{t|t} F_t^T P_{t+1|t}^{-1}, \quad (38)$$

for  $t = T, T-1, \dots, 1, 0$ . We desire to define a *smoothed credal set* of the form

$$\underline{X}_{t|T} = \{x \in \mathcal{R}^n: \|K_{t|T}^{-1}(x - c_{t|T})\| \leq 1\},$$

where  $K_{t|T}$ , the *smoothed credal matrix*, is suitably defined. To determine  $K_{t|T}$ , we determine linear relationship of the form

$$\hat{x}_{t|T} = A \hat{x}_{t|t} + b_t, \quad (39)$$

and apply Lemma 2. From (35) we obtain

$$[I - W_t H_t] F_{t-1} = K_{t|t} K_{t-1|t-1}^{-1}, \quad (40)$$

which, when substituted into the Kalman filter equations yields

$$\hat{x}_{t+1|t+1} = K_{t+1|t+1} K_{t|t}^{-1} \hat{x}_{t|t} + W_{t+1} z_{t+1}. \quad (41)$$

We motivate the development of the smoothed credal set by considering first the case  $t = T-1$ . Substituting (41) into

(36) with  $t = T-1$  gives

$$\begin{aligned} \hat{x}_{T-1|T} &= C_{T-1} [K_{T|T} K_{T-1|T-1}^{-1} \hat{x}_{T-1|T-1} + W_T z_T] \\ &\quad + (I - C_{T-1} F_{T-1}) \hat{x}_{T-1|T-1} \\ &= [C_{T-1} K_{T|T} K_{T-1|T-1}^{-1} + (I - C_{T-1} F_{T-1})] \\ &\quad \times \hat{x}_{T-1|T-1} + C_{T-1} W_T z_T. \end{aligned} \quad (42)$$

By defining

$$A = [C_{T-1} K_{T|T} K_{T-1|T-1}^{-1} + (I - C_{T-1} F_{T-1})] \quad (43)$$

$$b_{T-1} = C_{T-1} W_T z_T, \quad (44)$$

(42) becomes

$$\hat{x}_{T-1|T} = A \hat{x}_{T-1|T-1} + b_{T-1}. \quad (45)$$

Application of Lemma 2 with  $A$  and  $b$  defined by (43) and (44), respectively, and setting  $t = T-1$  yields

$$\begin{aligned} K_{T-1|T} &= A K_{T-1|T-1} \\ &= C_{T-1} K_{T|T} + [I - C_{T-1} F_{T-1}] K_{T-1|T-1}. \end{aligned}$$

To obtain the general smoother solution for  $t \leq T-2$ , assume that  $K_{t+1|T}$  and  $b_{t+1}$  have been calculated such that the following relationship holds for  $t+1$ :

$$\hat{x}_{t+1|T} = K_{t+1|T} K_{t+1|t+1}^{-1} \hat{x}_{t+1|t+1} + b_{t+1}. \quad (46)$$

Using this expression in (36) gives

$$\begin{aligned} \hat{x}_{t|T} &= C_t (K_{t+1|T} K_{t+1|t+1}^{-1} [K_{t+1|t+1} K_{t|t}^{-1} \hat{x}_{t|t} + W_{t+1} z_{t+1}] \\ &\quad + b_{t+1}) + (I - C_t F_t) \hat{x}_{t|t} \\ &= [C_t K_{t+1|T} K_{t|t}^{-1} + (I - C_t F_t)] \hat{x}_{t|t} + C_t b_{t+1} \\ &\quad + C_t K_{t+1|T} K_{t+1|t+1}^{-1} W_{t+1} z_{t+1}. \end{aligned} \quad (47)$$

Now define

$$b_t = C_t b_{t+1} + C_t K_{t+1|T} K_{t+1|t+1}^{-1} W_{t+1} z_{t+1}. \quad (48)$$

$$A = [C_t K_{t+1|T} K_{t|t}^{-1} + (I - C_t F_t)], \quad (49)$$

then (47) is in the form (39), thus we may invoke Lemma 2 as before to obtain

$$K_{t|T} = C_t K_{t+1|T} + (I - C_t F_t) K_{t|t}. \quad (50)$$

The derivation of the smoother is summarized in the following theorem.

**Theorem 4:** Assume the same hypotheses as for Theorem 3. Then the set-valued smoother is

$$\underline{X}_{t|T} = \{x \in \mathcal{R}^n: \|K_{t|T}^{-1}(x - c_{t|T})\| \leq 1\},$$

where

$$c_{t|T} = C_t c_{t+1|T} + (I - C_t F_t) c_{t|t}$$

and

$$K_{t|T} = C_t K_{t+1|T} + (I - C_t F_t) K_{t|t},$$



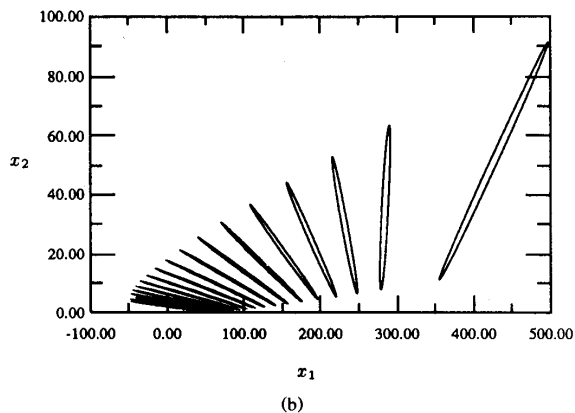
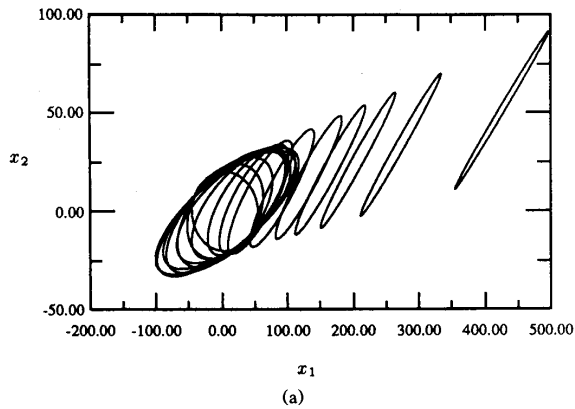


Fig. 1. Filtered and smoothed credal sets. (a) Filtered credal sets. (b) Smoothed credal sets.

with

$$P_{t|T} = P_{t|t} + C_t(P_{t+1|T} - P_{t+1|t})C_t^T,$$

$$C_t = P_{t|t}F_t^T P_{t+1|t}^{-1},$$

and  $P_{t|t}$ ,  $P_{t+1|t}$ ,  $K_{t|t}$ , and  $\underline{c}_{t|t}$  given by Theorem 3.

### VI. APPLICATION

Typical applications of estimation require point estimates. Not all estimation problems, however, require a point estimate at each time instant. For example, when the estimation of aircraft, missile, or submarine trajectories, there may not be the requirement for taking action at each sample time. With such tracking scenarios, especially if the observability of the target is in question (e.g., due to weak or fading signals, jamming, or the possibility of multiple targets), it would be wise to suspend judgment as long as possible, allow a set of trajectories to evolve and, only when required to do so, reduce the set estimate to a point estimate.

Even though a system may be uniformly observable and controllable, the convergence may be slow, and large transients may occur before the credal matrix ultimately begins

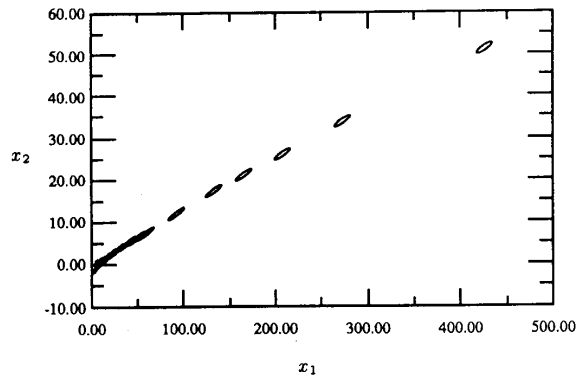


Fig. 2. Filtered constant-covariance contours.

to converge. Consider the following two-dimensional example. Let

$$F = \begin{bmatrix} 1.1 & 1 \\ 0 & 1.2 \end{bmatrix}, \quad G = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad H = [0.1 \ 0],$$

and suppose  $P_{0|0} = I$ ,  $Q = 0.001$ ,  $R = 2.0$ ,  $\underline{c}_{0|0} = [5 \ 0]$ , and  $K_{0|0} = \begin{bmatrix} 50 & 0 \\ 0 & 20 \end{bmatrix}$ . We assume all distributions are Gaussian, and use the Euclidean norm (32) for the definition of  $\underline{X}_{t|t}$ . Thus,

$$\underline{X}_{0|0} = \left\{ [x_1 - 5 \ x_2] \begin{bmatrix} 2500 & 0 \\ 0 & 400 \end{bmatrix}^{-1} \begin{bmatrix} x_1 - 5 \\ x_2 \end{bmatrix}^T \leq 1 \right\},$$

and the prior credal state generator consists of the set of Gaussian probability densities

$$\mathcal{V}_{0|0} = \{ \mathcal{N}(\underline{x}, I) : \underline{x} \in \underline{X}_{0|0} \}.$$

The initial credal set is an ellipse with semimajor and minor axes of length 50 and 20 units, respectively. We interpret this ellipse as being the family of all seriously possible initial conditions, i.e., the *a priori* distribution for the state is assumed to have a mean value lying in this region. Note that the *a priori* covariance is the identity matrix. Thus, the *a priori* distribution has a relatively small second moment compared to the size of the region of seriously possible first moments.

Figure 1 displays the evolution of a sequence of filtered and smoothed credal sets for this system. In these plots, the  $x_1$  element of the system is the abscissa, and the  $x_2$  element is the ordinate. Each elliptical credal set corresponds to a particular time; the trajectory of the system is moving up and to the right as time increases. For each time, each point in the corresponding ellipse is the output of a Kalman filter (Plot (a)) and fixed-interval smoother (Plot (b)) for a particular initial condition. Note that, as data are processed, the credal sets become more narrow in one direction and more elongated in the orthogonal direction. It is straightforward to establish the observability and controllability of this system, so it is known that the system is credible and the ellipsoids will ultimately converge to a point. For this example, however, it is seen that there is a significant transient before the credal sets begin to converge.

Figure 2 illustrates a sequence of filtered covariance contours defined by

$$(\underline{x} - \underline{c}_{t|t})^T P_{t|t}^{-1} (\underline{x} - \underline{c}_{t|t}) = 1,$$

where, as with the credal sets, the contour ellipses are centered at  $\underline{c}_{t|t}$ . A similar covariance ellipse is associated with each element of  $\underline{X}_{t|t}$ . The constant-covariance contour represents the second-moment of the conditional distribution, and admits an interpretation as the one-sigma ellipse of the estimate. It is a measure of the dispersion of the true state about a *single* estimated state, and is a valid measure of uncertainty only if the dispersion of first moments negligible. The credal sets, on the other hand, admit an interpretation as a measure of the dispersion of the first moments of the family of conditional distributions that make up the credal state at each time. The dispersion of the first moment, which occurs because of the lack of knowledge about the mean of the initial distribution, is a measure of dispersion *in addition to* that provided by the conditional second moment.

### VII. DISCUSSION

A classical way to deal with ignorance concerning the *a priori* distribution is to set the initial mean to some arbitrary value (typically, zero) and let the variance be extremely large (ideally approaching infinity), thereby deweighting completely the initial conditions. But such an approach is not completely satisfactory. There are often physical considerations that bound the values of the initial state, and these constraints are often more properly associated with bounds on the first moment, rather than on the second moment.

The additional measure of dispersion associated with the credal sets provides a means of factoring an agent's *ignorance* into the estimation problem in addition to specification of the variance. As discussed in [3] and elsewhere, one of the problems often associated with the Bayesian approach is that there is no way for an agent to differentiate between his ignorance (lack of knowledge) and his uncertainty (imprecise knowledge). By forcing the agent to specify one and only one prior distribution, the resulting estimate is represented as the unique best estimate, even though the specification of a unique prior may be made in an unwarranted and objectionable way (though required for operation of the estimator). By relaxing the requirement to specify a unique prior, the agent is permitted the opportunity to examine a family of estimates, each one associated with a specific prior, and thereby having the opportunity to assess the significance of the choice of prior.

A issue associated with set-valued estimation is that the estimate is *not* unique, and it may be unsettling to have to deal with a set of estimates. Unfortunately, this is the price to be paid if the agent is to acknowledge his true state of (or lack of) knowledge. If the agent is unable to reliably specify a unique prior distribution, then he should not expect to obtain a unique posterior distribution (and, hence, a unique estimate). On the fortunate side, however, if there are sufficient data available and the system is uniformly observable, then the set-valued estimator will asymptotically converge to a point estimate, which corresponds to the estimate that would be obtained for *any* choice of prior distribution in the *a priori* credal state.

For data-limited applications, however, even if the system is uniformly controllable and observable, it may be advisable to employ a set-valued estimator, thereby providing an agent with the opportunity of assessing the transient behavior of the credal set. For marginally observable systems (i.e., sys-

tems for which the sensitivity of the observables to changes in the state is extremely small) or partially observable systems (i.e., systems for which there is an unobservable subsystem), the use of point-valued estimates may prove to be inadequate. While the set-valued estimator does not yield "better" estimates in the sense of improving accuracy, it does provide the agent with a more complete picture of the sensitivity of the estimator to ignorance and uncertainty in the initial conditions.

### VIII. CONCLUSION

This paper extends the conventional notion of Bayes estimation as the calculation of a single conditional distribution to the calculation of a convex set of conditional distributions. It is shown that a Markov dynamics model and a white noise observation model yield an estimator structure that preserves convexity of the sets of distributions. It is also shown that a convex set of distributions may be represented by a set of distributions (e.g., the Gaussian probability densities), and that convexity is preserved in the estimator structure if only the generator set is considered. By taking the generator family of distributions to be Gaussian distributions with means in a convex set in state space and equal variances, a set-valued Kalman filter is developed that generates the equivalent of an infinity of Kalman filter, each with different initial conditions.

If a system is not uniformly observable, or if there are insufficient data to eliminate the effects of incorrect specification of the prior distribution of the system state, the choice of initial conditions may be critical for Bayesian estimation techniques such as the Kalman filter. In the interest of avoiding error, a set-valued estimator provides a set, or family, of all possible estimates that are consistent with the observed data and the contextual and logical information known about the system. The set-valued Kalman filter provides a means of propagating this trajectory credal set, rather than just one member of the set. By examining the structure of the credal set, it is possible to ascertain the effect of the initial conditions on the system state estimate.

### REFERENCES

- [1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, Ser. D: J. Basic Eng.*, vol. 82, pp. 35-45, 1960.
- [2] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York: Academic Press, 1970.
- [3] W. C. Stirling and D. R. Morrell, "Convex Bayes decision theory," *IEEE Trans. Syst., Man Cybern.*, vol. 21, pp. 173-183, Jan./Feb 1991.
- [4] I. Levi, *The Enterprise of Knowledge*. MIT Press, Cambridge, MA, 1980.
- [5] —, "Imprecision and indeterminacy in probability judgement," *Phil. Sci.*, vol. 52, no. 3, pp. 390-409, 1985.
- [6] D. R. Morrell, "A theory of set-valued estimation," Ph.D. thesis, Brigham Young University, Provo, UT, 1988.
- [7] D. R. Morrell and W. C. Stirling, "Set-valued filtering and smoothing," in *Conf. Record Twenty-Second Asilomar Conf. Signals Syst. Comput.*, pp. 236-239, Oct. 1988.
- [8] D. R. Morrell and W. C. Stirling, "Tracking of marginally observable targets," in *Proc. ICASSP89*, pp. 2720-2723, May 1989.
- [9] H. S. Witsenhausen, "A minimax control problem for sampled linear systems," *IEEE Trans. Automatic Contr.*, vol. AC-13, pp. 5-21, Feb. 1968.

- [10] \_\_\_\_\_, "Sets of possible states of linear systems given perturbed observations," *IEEE Trans. Automatic Contr.*, vol. AC-13, pp. 556-558, Oct. 1968.
- [11] F. C. Schweppe, "Recursive state estimation: Unknown but bounded errors and system inputs," *IEEE Trans. Automatic Contr.*, vol. AC-13, pp. 22-28, Feb. 1968.
- [12] D. P. Bertsekas and I. B. Rhodes, "Recursive state estimation for a set-membership description of uncertainty," *IEEE Trans. Automatic Contr.*, vol. AC-16, pp. 117-128, Apr. 1971.
- [13] A. J. Krener, "Kalman-Bucy and minimax filtering," *IEEE Trans. Automatic Contr.*, vol. AC-25, pp. 291-292, Apr. 1980.
- [14] I. Ya. Kats and A. B. Kurzhanskii, "Minimax multistep filtering in statistically indeterminate situations," *Avtomat. i Telemekh.*, vol. 11, p. 79, 1978.
- [15] B. N. Pshenichnyi and V. G. Pokotilo, "On observation problems in discrete systems," *PMM J. Applied Math. Mechanics*, vol. 45, no. 1, pp. 1-6, 1981.
- [16] B. I. Anan'ev and A. B. Kurzhanskii, "The nonlinear filtering problem for a multistage system with statistical uncertainty," *Proc. 2nd IFAC Symp.*, 1986, pp. 19-23.
- [17] Y. C. Ho and R. C. K. Lee, "A Bayesian approach to problems in stochastic estimation and control," *IEEE Trans. Automatic Contr.*, vol. AC-9, pp. 333-339, Oct. 1964.
- [18] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *ALAA J.*, vol. 3, pp. 1445-1450, 1965.

**Darryl R. Morrell** (S'85-M'88), for a photograph and biography, please see page 183 of this issue of the TRANSACTIONS.

**Wynn C. Stirling** (M'73), for a photograph and biography, please see page 183 of this issue of the TRANSACTIONS.