RESEARCH   **Open Access**

The interval versions of the Kalman filter and the EM algorithm

O Al-Gahtani[1*], J Al-Mutawa[2], M El-Gebeily[2] and R Agarwal[1,3]

[*]Correspondence:
obalgahtani@ksu.edu.sa
[1]Department of Mathematics, King
Saud University, Riyadh, Kingdom of
Saudi Arabia
Full list of author information is
available at the end of the article

Abstract

In this paper, we study state space models represented by interval parameters and noise. We introduce an interval version of the Expectation Maximization (EM) algorithm for the identification of the interval parameters of the system. We also introduce a suboptimal interval Kalman filter for the identification and estimation of the state vectors. The work requires the introduction of the concept of interval random variables which we also include in this work together with a study of their interval statistical properties such as expectation, conditional expectation and variance. Although the interval Kalman filter introduced here is suboptimal, it successfully recovers the state vectors to a high precision in the simulation examples we have run.

1 Introduction

In a state space model, some parameters of the system such as the coefficient matrices may not be precisely known or they gradually change with time. One way to account for these uncertainties is to allow such parameters to be represented by interval entities. The question then arises as to how to extend identification and estimation techniques to interval settings.

To our knowledge, no attempt has been made so far to extend identification techniques such as the EM algorithm to interval state space models. In this work, we give one such an extension.

In the existing literature, an optimal interval Kalman filter was attempted in [1]. That attempt suffered from the lack of proper definitions and rigorous treatment. The idea in [1] was to replace the interval system setting with the 'worst case inversion' while keeping everything else unchanged. So, the ultimate treatment in [1] amounts to the application of the traditional Kalman filter to the system representing the worst case scenario. This way the authors were able to avoid the difficulties that arise when dealing with interval arithmetic and concepts. On the other hand, this algorithm cannot be called optimal and the concept of the optimal interval Kalman filter remains an open question.

In our work, we introduce a spacial interval arithmetic that always produces results that are smaller (in the sense that it is contained) than the traditional interval arithmetic [2, 3]. This arithmetic enables the extension of the Kalman filter as well as the EM algorithm to interval setting in a true sense. In our restricted interval arithmetic, the interval Kalman filter we introduce here is optimal. However, with respect to the more general interval arithmetic, our interval Kalman filter is suboptimal.

Springer

2 A special interval arithmetic

We introduce a special set of interval operations that will enable the extension of the usual linear system concepts to the interval setting in a seamless manner. The more general definitions of the interval operations can be found in [2]. The arithmetic introduced here avoids such vague terms as 'interval extension', 'inclusion function', determinants *etc.* that have been used in the literature [1, 4–6].

All the interval operations adopted in this work stem from the view of an interval as a set of convex combinations of its endpoints:

$$I = [a, b] = \{x_\alpha = (1 - \alpha)a + \alpha b : \alpha \in [0, 1]\}.$$

Definition 1 Suppose $I = [a, b]$, $J = [c, d]$ are intervals and $\bullet \in \{+, -, *, \div\}$. Define the following interval operations:

$$I \bullet J = \{x_\alpha \bullet y_\alpha : \alpha \in [0, 1], x_\alpha \in I, y_\alpha \in J\},$$

with the usual restriction $0 \notin J$ if $\bullet = \div$.

Observe that all operations in Definition 1 result in intervals since they can be regarded as continuous functions defined on the unit interval $[0, 1]$. For example, a typical element in $I * J$ is $(1 - \alpha)^2 ac + \alpha(1 - \alpha)(ad + bc) + \alpha^2 bd$ which is a continuous function of α. The operations in Definition 1 give similar results to the usual interval operations as given in [2] when $\bullet \in \{+, -\}$, but generally they give only subintervals if $\bullet \in \{*, \div\}$. For example, if $I = [-2, 2]$, then $I * I = [0, 4]$ according to Definition 1, while the usual definition in [2] gives $I * I = [-4, 4]$.

The operations in Definition 1 are associative:

$$I + (J + K) = (I + J) + K,$$

$$I * (J * K) = (I * J) * K$$

and distributive:

$$I * (J + K) = I * J + I * K.$$

For example, distributivity is shown as follows:

$$
\begin{aligned}
I * (J + K) &= \{I_\alpha (J + K)_\alpha : \alpha \in [0, 1]\} \\
&= \{I_\alpha (J_\alpha + K_\alpha) : \alpha \in [0, 1]\} \\
&= \{I_\alpha J_\alpha + I_\alpha K_\alpha : \alpha \in [0, 1]\} \\
&= \{(IJ)_\alpha + (IK)_\alpha : \alpha \in [0, 1]\} \\
&= \{(IJ + IK)_\alpha : \alpha \in [0, 1]\} \\
&= IJ + IK.
\end{aligned}
$$

These two properties, which are missing in the usual interval operations, will enable the extension of many results from usual state space models to interval state space models. On

the other hand, these definitions were motivated by our attempt to arrive at a definition of interval random variables and investigate the corresponding statistical properties. We feel that they are the natural ones to handle interval systems. This feeling is reassured by the numerical results we obtained in the simulation examples (see Section 5). While we expected to obtain a construction of a suboptimal interval Kalman filter, the constructed filter was actually able to recover the exact simulated intervals rather than subintervals.

Interval vectors and matrices are defined similarly:

- A vector $\mathbf{v} \in \mathbf{IR}^n$ is defined as

$$\mathbf{v} = [\mathbf{a}, \mathbf{b}] = \left\{ \mathbf{x}_\alpha = (1 - \alpha)\mathbf{a} + \alpha\mathbf{b} : \alpha \in [0, 1] \right\},$$

where

$$\mathbf{a}, \mathbf{b} \in \mathbb{R}^n, \qquad \mathbf{a} \le \mathbf{b},$$

and the inequality holds componentwise.

- A matrix $\mathbf{A} \in \mathbf{IR}^{n \times n}$ is defined as

$$\mathbf{A} = [A, B] = \left\{ X_\alpha = (1 - \alpha)A + \alpha B : \alpha \in [0, 1] \right\},$$

where

$$A, B \in \mathbb{R}^{n \times n}, \qquad A \le B,$$

and the inequality holds componentwise.

Definition 2 Given a function $f : \mathbb{R}^k \to \mathbb{R}^n$ and an interval vector $\mathbf{v} \in \mathbf{IR}^k$, we define

$$f(\mathbf{v}) = \left\{ f(\mathbf{x}_\alpha) : \alpha \in [0, 1] \right\}.$$

If f is continuous, then $f(\mathbf{v})$ is an interval vector. All operations on functions are extended to interval settings in the same way. For example,

$$fg(\mathbf{v}) = f(\mathbf{v})g(\mathbf{v}) = \left\{ f(\mathbf{x}_\alpha)g(\mathbf{x}_\alpha) : \alpha \in [0, 1] \right\},$$

$$\frac{df}{d\mathbf{v}} = \left\{ \frac{df}{d\mathbf{x}_\alpha} : \alpha \in [0, 1] \right\},$$

provided that the involved operations make sense. In the same spirit, interval matrix operations are defined as follows:

- The interval determinant is defined by

$$\det(\mathbf{A}) = \left\{ \det(X_\alpha) : \alpha \in [0, 1] \right\}.$$

- The interval adjoint is defined by

$$\mathrm{adj}(\mathbf{A}) = \left\{ \mathrm{adj}(X_\alpha) : \alpha \in [0, 1] \right\}.$$

- The interval inverse is defined by

$$
\begin{aligned}
\mathbf{A}^{-1} &= \left\{ X_\alpha^{-1} : \alpha \in [0,1] \right\} \\
&= \left\{ \frac{\mathrm{adj}(X_\alpha)}{\det(X_\alpha)} : \alpha \in [0,1] \right\} \\
&= \frac{\mathrm{adj}(\mathbf{A})}{\det(\mathbf{A})}.
\end{aligned}
$$

The continuous dependence of the determinant of a matrix on the elements of the matrix implies that all the above operations produce interval entities. Naturally, these special definitions produce results that are contained in the corresponding usual definitions. To give an example, we use the definition of the inverse interval matrix $\mathbf{A}^{-1}$ according to [3]:

$$
\mathbf{A}^{-1} = \left[\left\{ X^{-1} : X \in \mathbf{A} \right\} \right],
$$

where $[\mathcal{S}]$ is the smallest interval vector (matrix) containing $\mathcal{S}$. If

$$
\mathbf{A} = \begin{bmatrix} [2] & [-1,1] \\ [-1,1] & [2] \end{bmatrix},
$$

then

$$
\mathcal{S} = \left\{ X^{-1} : X \in \mathbf{A} \right\} = \left\{ \begin{bmatrix} \frac{2}{4-rs} & -\frac{r}{4-rs} \\ -\frac{r}{4-rs} & \frac{2}{4-rs} \end{bmatrix} : r,s \in [-1,1] \right\}.
$$

One can show that the set of points in $\mathbb{R}^2$ with coordinates equal to the first row of the elements of $\mathcal{S}$ forms a polygonal (non-rectangular) region with vertices at $(\frac{2}{5}, \frac{1}{5})$, $(\frac{1}{2}, 0)$, $(\frac{2}{3}, \frac{1}{3})$, $(\frac{2}{5}, -\frac{1}{5})$, $(\frac{2}{3}, -\frac{1}{3})$. Thus,

$$
\mathbf{A}^{-1} = [\mathcal{S}] = \begin{bmatrix} [\frac{2}{5}, \frac{2}{3}] & [-\frac{1}{3}, \frac{1}{3}] \\ [-\frac{1}{3}, \frac{1}{3}] & [\frac{2}{5}, \frac{2}{3}] \end{bmatrix}.
$$

The inverse in our sense is

$$
\mathbf{A}^{-1} = \begin{bmatrix} [\frac{1}{2}, \frac{2}{3}] & [0, \frac{1}{3}] \\ [0, \frac{1}{3}] & [\frac{1}{2}, \frac{2}{3}] \end{bmatrix}.
$$

- Suppose that $\mathbf{A}^{-1}$ exists. We define the solution of the interval linear system $\mathbf{AX} = \mathbf{b}$ to be

$$
\mathbf{X} = \left\{ X \in \mathbb{R}^n : A_\alpha X = b_\alpha, \alpha \in [0,1] \right\}.
$$

Clearly, X is an interval vector. The usual definition is

$$
\mathcal{S} = \left\{ X \in \mathbb{R}^n : \exists A \in \mathbf{A}, b \in \mathbf{b}, AX = b \right\}.
$$

Obviously, our definition produces a smaller interval vector. In fact, if $\mathbf{A}^{-1}$ exists in the sense of [3], then

$$\mathbf{X} \subset \mathcal{S} \subset [\mathcal{S}] \subset \mathbf{A}^{-1}\mathbf{b}.$$

The last inclusion holds because if $X \in \mathcal{S}$, then there is an $A \in \mathbf{A}$ and a $b \in \mathbf{b}$ with $AX = b$. Then $X = A^{-1}b \in \mathbf{A}^{-1}b \subset \mathbf{A}^{-1}\mathbf{b}$. Thus, $\mathcal{S} \subset \mathbf{A}^{-1}\mathbf{b}$. Noting that $\mathbf{A}^{-1}\mathbf{b}$ is an interval vector and $[\mathcal{S}]$ is minimal, we get that $[\mathcal{S}] \subset \mathbf{A}^{-1}\mathbf{b}$.

For the rest of this paper, we will use the special interval operations defined above.

Finally, for error estimates, we need to introduce the distance between two intervals $I = [a, b], J = [c, d]$. This is defined by

$$q(I, J) := \max\{|a - c|, |b - d|\}.$$

The map q defines a metric in **IR**.

3 Interval random variables

We begin by discussing the measurability of set-valued maps and then introduce the definition of an interval random variable. The basic definitions and more details can be found in [7]. A measurable space $(\Omega, \mathcal{A})$ consists of a basic set Ω together with a σ-algebra $\mathcal{A}$ of subsets of Ω called measurable sets. Here, we consider closed convex value set-valued maps $F : \Omega \rightrightarrows \mathbb{R}^k$, *i.e.*, $F(\omega)$ is a closed convex subset of $\mathbb{R}^k$ for each $\omega \in \Omega$. This is the case when F is interval valued. The latter notion means that for each $\omega \in \Omega$, the components of $F(\omega)$ are closed intervals in $\mathbb{R}$.

We first define what it means for a set-valued map to be measurable. Recall that the inverse image of a set $S \subset \mathbb{R}^k$ under the set-valued map F is defined by

$$F^{-1}(S) = \{\omega \in \Omega : F(\omega) \cap S \neq \varnothing\},$$

and that the graph of F (denoted by G_F) is defined by

$$G_F = \{(\omega, y) : \omega \in \Omega, y \in F(\omega)\}.$$

Definition 3 Let $(\Omega, \mathcal{A})$ be a measurable space and $F : \Omega \rightrightarrows \mathbb{R}^k$ be a set-valued map. F is called measurable if the inverse image of each open set is a measurable set: if $O \subset \mathbb{R}^k$ is open, then $F^{-1}(O) \in \mathcal{A}$.

We are now in a position to introduce the definition of interval random variables and interval stochastic processes.

Definition 4 Let $(\Omega, \mathcal{S}, P)$ be a probability space. An interval-valued map $X : \Omega \rightrightarrows \mathbb{R}^k$ is called an interval random variable if

1. X is measurable, and
2. the function $x \mapsto p_x$ is continuous on X, where p_x is the probability density function for the random variable x.

An interval stochastic process is an indexed set of interval random variables.

The probability density function p_X is then the interval-valued function

$$p_X = \{p_x : x \in X\}.$$

In order to study the expectations and variances of interval random variables, we need to discuss first the integral of set-valued maps and, in particular, interval-valued maps. The discussion begins with the notion of measurable selections.

Definition 5 Let $(\Omega, \mathcal{A})$ be a measurable space and $F : \Omega \rightrightarrows \mathbb{R}^k$ be a measurable set-valued map. A measurable selection of F is a measurable map $f : \Omega \to \mathbb{R}^k$ satisfying $f(\omega) \in F(\omega)$ for each $\omega \in \Omega$.

It is well known that every measurable set-valued map has at least one measurable selection [8]. Furthermore, we have the following equivalences [7].

Theorem 6 *Let $(\Omega, \mathcal{A})$ be a measurable space and denote by $\mathcal{B}$ the σ-algebra of Borel sets in $\mathbb{R}^k$. Let $F : \Omega \rightrightarrows \mathbb{R}^k$ be a set-valued map. The following are equivalent.*

1. *F is measurable.*
2. *$G_F \in \mathcal{A} \otimes \mathcal{B}$.*
3. *$F^{-1}(B) \in \mathcal{A}$ for every $B \in \mathcal{B}$.*
4. *There exists a sequence of measurable selections $\{f_n\}_{n=1}^{\infty}$ of F such that*

$$F(\omega) = \overline{\bigcup_{n \geq 1} f_n(\omega)}$$

for each $\omega \in \Omega$.

A countable family of measurable selections satisfying the last property is called *dense*.

Let $F : \Omega \rightrightarrows \mathbb{R}^k$ be an interval-valued map. We define the two special functions l_F and r_F such that $l_F(\omega) = a(\omega)$ and $r_F(\omega) = b(\omega)$, where $F(\omega) = [a(\omega), b(\omega)]$ for each $\omega \in \Omega$. The next lemma shows that l_F and r_F are measurable selections of F when the latter is measurable.

Lemma 7 *Let $F : \Omega \rightrightarrows \mathbb{R}^k$ be a measurable interval-valued map. Then the point functions l_F and r_F are measurable selections of F.*

Proof Choose a sequence of measurable selections $\{f_n\}_{n=1}^{\infty}$ of F such that

$$F(\omega) = \overline{\bigcup_{n \geq 1} f_n(\omega)}.$$

Then $l_F(\omega) = \inf_{n \geq 1} f_n(\omega)$ and $r_F(\omega) = \sup_{n \geq 1} f_n(\omega)$ (here the inf and sup operations are taken componentwise). Since the inf and the sup operators preserve measurability, we see that the functions l_F and r_F are measurable selections of F. □

Example Let $\Omega = [1, \infty)$ and define $F : \Omega \rightrightarrows \mathbb{R}$ by

$$F(t) = \left[t, t + \frac{1}{t}\right].$$

Let $\{r_n\}_{n=1}^{\infty}$ be an enumeration of the rational numbers in the interval $[0,1]$, and let us assume that $r_1 = 1$, $r_2 = 0$. Define $f_n : [1,\infty) \to \mathbb{R}$ by

$$f_n(t) = r_n t + (1 - r_n)\left(t + \frac{1}{t}\right).$$

Thus, $l_F(t) = t = f_1(t)$ and $r_F(t) = (t + \frac{1}{t}) = f_2(t)$. For every $t \in [1,\infty)$, the set $\{r_n t + (1-r_n)(t + \frac{1}{t})\}_{n=1}^{\infty}$ is dense in the interval $[t, t + \frac{1}{t}]$. Thus, $F(t) = \overline{\bigcup_{n\geq 1} f_n(t)}$.

Now suppose that $(\Omega, \mathcal{A}, \mu)$ is a measure space and $F : \Omega \rightrightarrows \mathbb{R}^k$ is a set-valued map. A measurable selection f of F is an integrable selection if f is integrable with respect to the measure μ. The set of all integrable selections of F will be denoted by $\mathcal{F}$. The map F is called integrably bounded if there exists a μ-integrable function $g \in L^1(\Omega; \mathbb{R}, \mu)$ such that $F(\omega) \subset g(\omega)\mathbf{B}$ for μ-almost every $\omega \in \Omega$. Here, $\mathbf{B}$ denotes the unit ball in $\mathbb{R}^k$. In this case, every measurable selection f of F is also an integrable selection since $f(\omega) \in F(\omega) \subset g(\omega)\mathbf{B}$ implies that $\|f(\omega)\| \leq |g(\omega)|$, where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^k$.

Definition 8 The integral of a set-valued map F is defined to be the set of integrals of integrable selections of F. That is,

$$\int_{\Omega} F \, d\mu = \left\{ \int_{\Omega} f \, d\mu : f \in \mathcal{F} \right\}. \tag{1}$$

We shall say that F is integrable if every measurable selection is integrable.

We have the following immediate properties:

$$\int_{\Omega} \lambda F \, d\mu = \lambda \int_{\Omega} F \, d\mu, \tag{2}$$

$$\overline{\int_{\Omega} (F_1 + F_2) \, d\mu} = \overline{\int_{\Omega} F_1 \, d\mu} + \overline{\int_{\Omega} F_2 \, d\mu}. \tag{3}$$

Lemma 9 *Let $F : \Omega \rightrightarrows \mathbb{R}^k$ be an interval-valued map. If l_F and r_F are integrable, then F is integrable and*

$$\int_{\Omega} F \, d\mu = \left[\int_{\Omega} l_F \, d\mu, \int_{\Omega} r_F \, d\mu \right]$$

$$= \left\{ \int_{\Omega} f_\alpha \, d\mu : f_\alpha = \alpha l_F + (1 - \alpha) r_F, \alpha \in [0,1] \right\}.$$

Proof The first equality is shown as follows. Since for every $\omega \in \Omega$ and every integrable selection f of F we have $l_F(\omega) \leq f(\omega) \leq r_F(\omega)$,

$$\int_{\Omega} l_F(\omega) \, d\mu \leq \int_{\Omega} f(\omega) \, d\mu \leq \int_{\Omega} r_F(\omega) \, d\mu.$$

Therefore,

$$\int_{\Omega} F \, d\mu \subseteq \left[\int_{\Omega} l_F \, d\mu, \int_{\Omega} r_F \, d\mu \right].$$

On the other hand, let $\theta \in [\int_\Omega l_F \, d\mu, \int_\Omega r_F \, d\mu]$. We may write $\theta = (1 - \alpha) \int_\Omega l_F \, d\mu + \alpha \int_\Omega r_F \, d\mu$ for some $\alpha \in [0, 1]$. Then

$$\theta = \int_\Omega \big((1 - \alpha)l_F + \alpha r_F\big) \, d\mu$$

$$= \int_\Omega f_\alpha \, d\mu,$$

where $f_\alpha = (1 - \alpha)l_F + \alpha r_F$. Hence, $\theta \in \int_\Omega F \, d\mu$.

The second equality is an immediate consequence of this. □

It will always be assumed that both l_F and r_F are integrable.

Example Let Ω and F be defined as in the previous example. Let μ be the measure defined by

$$d\mu = \frac{1}{t^3} \, dt.$$

Then

$$\int_\Omega F \, d\mu = \left[\int_1^\infty l_F(t) \, d\mu, \int_1^\infty r_F(t) \, d\mu \right] = \left[1, \frac{4}{3} \right].$$

In view of (3), we have the following corollary.

Corollary 10 *Let $F_1, F_2 : \Omega \rightrightarrows \mathbb{R}^k$ be integrable interval-valued maps. Then*

$$\int_\Omega (F_1 + F_2) \, d\mu = \int_\Omega F_1 \, d\mu + \int_\Omega F_2 \, d\mu$$

$$= \left[\int_\Omega l_{F_1} \, d\mu, \int_\Omega r_{F_1} \, d\mu \right] + \left[\int_\Omega l_{F_2} \, d\mu, \int_\Omega r_{F_2} \, d\mu \right]$$

$$= \left[\int_\Omega (l_{F_1} + l_{F_2}) \, d\mu, \int_\Omega (r_{F_1} + r_{F_2}) \, d\mu \right].$$

Let $(\Omega, \mathcal{S}, P)$ be a probability space, and let $Z : \Omega \rightrightarrows \mathbb{R}^k$ be an interval random variable. We have

$$Z(\omega) = \big[l_Z(\omega), r_Z(\omega) \big] = \big\{ z_\alpha := (1 - \alpha)l_Z(\omega) + \alpha r_Z(\omega) : \alpha \in [0, 1] \big\}.$$

We shall say that Z is *normally distributed* if each $z \in Z$ is normally distributed. An interval stochastic process $\{Z_t\}_{t \in T}$ will be called normally distributed if for each $t \in T$, Z_t is normally distributed.

Let Z be an interval random variable. Then for each $z \in Z$,

$$p_{l_z} \leq p_z \leq p_{r_z}.$$

By the continuity of $z \mapsto p_z$,

$$p_Z = [p_{l_z}, p_{r_z}].$$

This means that

$$l_{p_Z} = p_{l_z}, \qquad r_{p_Z} = p_{r_z}.$$

Guided by this and Lemma 9, we can define the *interval expectation* of the interval random variable Z as follows.

Definition 11 The interval expectation of an interval random variable Z is defined as

$$E(Z) = \big[E(l_Z), E(r_Z)\big].$$

This definition coincides with Definition 2 since

$$
\begin{aligned}
\big[E(l_Z), E(r_Z)\big] &= \big\{\alpha E(l_Z) + (1-\alpha)E(r_Z) : \alpha \in [0,1]\big\} \\
&= \big\{E\big(\alpha l_Z + (1-\alpha)r_Z\big) : \alpha \in [0,1]\big\} \\
&= \big\{E(z_\alpha) : \alpha \in [0,1]\big\}.
\end{aligned}
$$

It should also be noted that the expectation of a vector random variable is the vector of expectations of its components.

It follows from equations (2) and (3) that

$$E(\lambda Z) = \lambda E(Z),$$

$$E(Z_1 + Z_2) = E(Z_1) + E(Z_2).$$

Also, if $I = [a, b]$ and Z is an interval random variable, then

$$
\begin{aligned}
E(IZ) &= \big\{E(t_\alpha z_\alpha) : \alpha \in [0,1]\big\} \\
&= \big\{t_\alpha E(z_\alpha) : \alpha \in [0,1]\big\} \\
&= I * \big\{E(z_\alpha) : \alpha \in [0,1]\big\} \\
&= IE(Z).
\end{aligned}
$$

The same is true if **I** is an interval vector and **Z** is an interval random variable.

More generally, if **A** is a $k \times k$ interval matrix and if its columns are denoted by the interval vectors $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$, then

$$
\begin{aligned}
E(\mathbf{AZ}) &= E\left(\sum_{j=1}^{k} \mathbf{A}_i Z_j\right) \\
&= \sum_{j=1}^{k} E(\mathbf{A}_i Z_j) = \sum_{j=1}^{k} \mathbf{A}_i E(Z_j) \\
&= \mathbf{A}E(\mathbf{Z}).
\end{aligned}
$$

To introduce covariance of two interval random variables Y, Z, we need to assume that the function $(x, y) \mapsto p_{x,y}$ is continuous on $Y \times Z$. Here, $p_{x,y}$ is the joint probability density function of the two random variables x, y.

Definition 12 The interval covariance of two interval random variables Y, Z is defined as

$$\mathrm{Cov}(Y, Z) = \big\{\mathrm{Cov}(y_\alpha, z_\alpha) : \alpha \in [0, 1]\big\}.$$

To see that $\mathrm{Cov}(Y, Z)$ is an interval, note that

$$\mathrm{Cov}(Y, Z) = \big\{\mathrm{Cov}\big((1 - \alpha)l_Y + \alpha r_Y, (1 - \alpha)l_Z + \alpha r_Z\big) : \alpha \in [0, 1]\big\}$$
$$= \big\{(1 - \alpha)^2 \mathrm{Cov}(l_Y, l_Z) + \alpha(1 - \alpha)\mathrm{Cov}(l_Y, r_Z)$$
$$+ \alpha(1 - \alpha)\mathrm{Cov}(r_Y, l_Z) + \alpha^2 \mathrm{Cov}(r_Y, r_Z) : \alpha \in [0, 1]\big\}.$$

If $Y = Z$, we get the definition of the variance of an interval random variable Z as

$$\mathrm{Var}(Z) = \big\{\mathrm{Var}(z_\alpha) : \alpha \in [0, 1]\big\}$$
$$= \big\{(1 - \alpha)^2 \mathrm{Var}(l_Z) + 2\alpha(1 - \alpha)\mathrm{Cov}(l_Z, r_Z) + \alpha^2 \mathrm{Var}(r_Z) : \alpha \in [0, 1]\big\}$$

which is also an interval. Elementary calculus considerations reveal that

$$\mathrm{Var}(Z) = \left[\frac{ab - c^2}{a + b - 2c}, \max\{a, b\}\right],$$

where $a = \mathrm{Var}(l_Z)$, $b = \mathrm{Var}(r_Z)$, $c = \mathrm{Cov}(l_Z, r_Z)$. This last equation provides a formula for computing the interval $\mathrm{Var}(Z)$.

For interval random vectors, the above definitions hold componentwise.

The two interval random variables Y, Z will be called uncorrelated if for each $y_\alpha \in Y$, $z_\alpha \in Z$, y_α, z_α are uncorrelated. Therefore, Y, Z are uncorrelated if and only if $\mathrm{Cov}(Y, Z) = [0]$.

It is now straightforward to check the following theorem.

Theorem 13 *Let* $\mathbf{Y}, \mathbf{Z} \in \mathbf{IR}^k$, $\mathbf{W} \in \mathbf{IR}^m$, $\mathbf{IR}^n$ *be interval random vectors, and let* $\mathbf{A} \in \mathbf{IR}^{k' \times k}$, $\mathbf{B} \in \mathbf{IR}^{m' \times m}$, $\lambda \in \mathbf{IR}$, *then*

1. $\mathrm{Cov}(\lambda \mathbf{Y}, \mathbf{W}) = \lambda \mathrm{Cov}(\mathbf{Y}, \mathbf{W})$,
2. $\mathrm{Cov}(\mathbf{Y} + \mathbf{Z}, \mathbf{W}) = \mathrm{Cov}(\mathbf{Y}, \mathbf{W}) + \mathrm{Cov}(\mathbf{Z}, \mathbf{W})$,
3. $\mathrm{Cov}(\mathbf{A}\mathbf{Y}, \mathbf{B}\mathbf{W}) = \mathbf{A}\,\mathrm{Cov}(\mathbf{Y}, \mathbf{W})\mathbf{B}^T$.

The assumed continuous dependence of the probability density function (joint density function) on the random variable (variables) in an interval random variable (interval random variables) implies that the conditional probability density function is also continuous. This guarantees that the generalization of the conditional density function to the interval setting is always an interval.

Definition 14 The interval conditional expectation is defined as

$$E(Z|Y) = \big\{E(z_\alpha|y_\alpha) : \alpha \in [0, 1]\big\}$$
$$= \big\{\alpha E(l_Z|y_\alpha) + (1 - \alpha)E(r_Z|y_\alpha) : \alpha \in [0, 1]\big\}.$$

The following theorem is easily checked.

Theorem 15 *For vector random variables* $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ *and interval matrix* $\mathbf{A}$ *of appropriate dimensions*,

1. $E(\mathbf{X} + \mathbf{Y}|\mathbf{Z}) = E(\mathbf{X}|\mathbf{Y}) + E(\mathbf{Y}|\mathbf{Z})$,
2. $E(\mathbf{A}\mathbf{Y}|\mathbf{Z}) = \mathbf{A}E(\mathbf{Y}|\mathbf{Z})$.

4 The interval state space model

The interval state space model we will consider here is one of the form

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{w}_t, \tag{4}$$

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{v}_t, \quad t \geq 0, \tag{5}$$

where $\mathbf{A} \in \mathbf{IR}^{k \times k}$, $\mathbf{H} \in \mathbf{IR}^{p \times k}$ are interval matrices and $\mathbf{w}_t \in \mathbf{IR}^k$, $\mathbf{v}_t \in \mathbf{IR}^p$ are zero-mean Gaussian white-noise interval processes, with

$$\mathrm{Cov}\left(\begin{bmatrix} \mathbf{w}_t \\ \mathbf{v}_t \end{bmatrix}, \begin{bmatrix} \mathbf{w}_s \\ \mathbf{v}_s \end{bmatrix}\right) = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \delta_{ts},$$

while the initial state $\mathbf{x}_0$ is assumed to be an interval random variable having zero-mean, interval variance matrix $\mathbf{\Pi}_0$ and to be uncorrelated to $\{\mathbf{w}_t\}$ and $\{\mathbf{v}_t\}$ for all $t \geq 0$. The matrices $\mathbf{Q} \in \mathbf{IR}^{k \times k}$, $\mathbf{R} \in \mathbf{IR}^{p \times p}$ are also allowed to be interval matrices. For the time being, we assume that the matrices $\mathbf{F}$, $\mathbf{H}$, $\mathbf{Q}$, $\mathbf{R}$ are known *a priori*. We thus have the properties

$$\mathrm{Cov}(\mathbf{w}_t, \mathbf{x}_s) = \mathbf{0}, \qquad \mathrm{Cov}(\mathbf{v}_t, \mathbf{x}_s) = \mathbf{0}, \quad s \leq t,$$

$$\mathrm{Cov}(\mathbf{w}_t, \mathbf{y}_t) = \mathbf{Q}, \qquad \mathrm{Cov}(\mathbf{v}_t, \mathbf{y}_t) = \mathbf{R}.$$

For the state covariance matrix

$$\mathbf{\Pi}_t = \mathrm{Cov}(\mathbf{x}_t, \mathbf{x}_t),$$

we have the recursion

$$\mathbf{\Pi}_{t+1} = \mathbf{A}\mathbf{\Pi}_t\mathbf{A}^T + \mathbf{Q}, \quad t \geq 0$$

with initial value $\mathbf{\Pi}_0$.

4.1 The interval Kalman filter

In this section, we give a summary of the interval settings of the Kalman filter and the EM algorithm. The ground work that we did in the previous two section should reveal that it is possible to apply both methods to interval state space models. Let $\mathbf{Y}_s = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_s\}$ be a sequence of interval measurements up to time s and let

$$\mathbf{x}_t^s = E(\mathbf{x}_t|\mathbf{Y}_s).$$

The expectation is a forecast for $s < t$, a filtered value for $s = t$ and a smoothed value for $s > t$. The least square estimation error is defined by

$$\mathbf{P}_t^s = E\big((\mathbf{x}_t - \mathbf{x}_t^s)(\mathbf{x}_t - \mathbf{x}_t^s)^T\big).$$

The interval Kalman filter is defined in two main steps: estimation and forecast [9, 10]. The estimation step is given by

$$\mathbf{x}_t^{t-1} = \mathbf{A}\mathbf{x}_{t-1}^{t-1}, \qquad \mathbf{P}_t^{t-1} = \mathbf{A}\mathbf{P}_{t-1}^{t-1}\mathbf{A}^T + \mathbf{Q},$$

and the forecast step is given by

$$\mathbf{x}_t^t = \mathbf{x}_{t-1}^{t-1} + \mathbf{K}_t\big(\mathbf{y}_t - \mathbf{H}\mathbf{x}_{t-1}^{t-1}\big), \tag{6}$$

$$\mathbf{P}_t^t = (\mathbf{I} - \mathbf{K}_t\mathbf{H})\mathbf{P}_t^{t-1}, \quad t = 1, 2, \ldots, n, \tag{7}$$

where

$$\mathbf{K}_t = \mathbf{P}_t^{t-1}\big(\mathbf{H}\mathbf{P}_t^{t-1}\mathbf{H}^T + \mathbf{R}\big)^{-1}$$

is called the Kalman gain. The initial conditions are $\mathbf{x}_0^0 = \boldsymbol{\mu}_0$ and $\mathbf{P}_0^0 = \boldsymbol{\Sigma}_0$.

The interval Kalman smoother, which is needed for the EM algorithm, is defined by

$$\mathbf{x}_t^n = \mathbf{x}_t^t + \mathbf{J}_t\big(\mathbf{x}_{t+1}^n - \mathbf{x}_{t+1}^t\big), \tag{8}$$

$$\mathbf{P}_t^n = \mathbf{P}_t^t + \mathbf{J}_t\big(\mathbf{P}_{t+1}^n - \mathbf{P}_{t+1}^t\big)\mathbf{J}_t^T, \tag{9}$$

where

$$\mathbf{J}_t = \mathbf{P}_t^t\mathbf{A}^T\big(\mathbf{P}_{t+1}^t\big)^{-1}.$$

The initial conditions $\mathbf{x}_n^n$ and $\mathbf{P}_n^n$ in this case are found from (6) and (7). The EM algorithm also needs the so-called lag-one covariance smoother defined by

$$\mathbf{P}_{n,n-1}^n = (\mathbf{I} - \mathbf{K}_n\mathbf{H})\mathbf{A}\mathbf{P}_{n-1}^{n-1},$$
$$\mathbf{P}_{t,t-1}^n = \mathbf{P}_t^t\mathbf{J}_{t-1}^T + \mathbf{J}_t^T\big(\mathbf{P}_{t+1,t}^n - \mathbf{A}\mathbf{P}_t^t\big)\mathbf{J}_{t-1}^T, \quad t = n-1, n-2, \ldots, 1. \tag{10}$$

4.1.1 The EM algorithm in interval setting

The EM algorithm [11–14] tries to estimate the parameter set $\boldsymbol{\Theta} = \{\mathbf{A}, \mathbf{H}, \mathbf{Q}, \mathbf{R}\}$ of the system (4), (5) by maximizing the likelihood of its probability density function $P(\mathbf{x}, \boldsymbol{\Theta})$. The algorithm consists of two steps, the expectation step (E-step) and the maximization step (M-step). The E-step uses the current available parameter set $\boldsymbol{\Theta}$ to obtain estimates for the state vector and the least square error. This step is based on the Kalman filter and Kalman smoother. The M-step uses the current estimated values of the state vector and errors to obtain a new parameter set according to the equations

$$\mathbf{A} = \mathbf{C}\mathbf{B}^{-1}, \tag{11}$$

$$\mathbf{Q} = \frac{1}{n}\big(\mathbf{D} - \mathbf{C}\mathbf{B}^{-1}\mathbf{C}^T\big), \tag{12}$$

$$\mathbf{H} = \mathbf{SD}^{-1}, \tag{13}$$

$$\mathbf{R} = \frac{1}{n}\left(\mathbf{U} - \mathbf{SD}^{-1}\mathbf{S}^T\right), \tag{14}$$

where

$$\mathbf{B} = \sum_{t=1}^{n} \mathbf{x}_{t-1}^n \left(\mathbf{x}_{t-1}^n\right)^T + \mathbf{P}_{t-1}^n, \qquad \mathbf{C} = \sum_{t=1}^{n} \mathbf{x}_t^n \left(\mathbf{x}_{t-1}^n\right)^T + \mathbf{P}_{t,t-1}^n,$$

$$\mathbf{D} = \sum_{t=1}^{n} \mathbf{x}_t^n \left(\mathbf{x}_t^n\right)^T + \mathbf{P}_t^n, \qquad \mathbf{S} = \sum_{t=1}^{n} \mathbf{y}_t \left(\mathbf{x}_t^n\right)^T,$$

$$\mathbf{U} = \sum_{t=1}^{n} \mathbf{y}_t \mathbf{y}_t^T.$$

The method can be summarized as follows.

1. Initialize the procedure by selecting starting values for the elements of the parameter set $\mathbf{\Theta}^{(0)} = \{\mathbf{A}^{(0)}, \mathbf{H}^{(0)}, \mathbf{Q}^{(0)}, \mathbf{R}^{(0)}\}$ and estimate $\boldsymbol{\mu}_0$.
2. (E-step) For $j = 1, 2, \ldots$, use the parameter set $\mathbf{\Theta}^{(j-1)}$ to estimate the smoothed values $\mathbf{x}_t^n, \mathbf{P}_t^n, \mathbf{P}_{t,t-1}^n$ (equations (8)-(10)) for $t = 1, 2, \ldots, n$.
3. (M-step) Calculate a new set of parameters $\mathbf{\Theta}^{(j)}$ using equations (11)-(14).
4. Repeat steps 2 and 3 above until convergence is achieved.

5 Simulation results

A 500 Monte Carlo simulation is performed to illustrate the utility of the interval EM algorithm estimate. The observed data are generated according to the second order interval state space model

$$x_{t+1} = \begin{bmatrix} [0.3, 0.5] & [-0.05, 0.25] \\ -0.1 & [-0.05, 0.45] \end{bmatrix} x_t + w_t, \quad t = 1, \ldots, 1{,}000, \tag{15}$$

$$y_t = \begin{bmatrix} 0 & [0.9, 1.1] \end{bmatrix} x_t + v_t, \tag{16}$$

where w_t and v_t are independent identically distributed (i.i.d.) Gaussian noises such that

$$w_t \sim \mathrm{N}\left(0, \begin{bmatrix} [0.05, 0.15] & 0 \\ 0 & [0.05, 0.15] \end{bmatrix}\right),$$

$$R_t \sim \mathrm{N}\left(0, [0.09, 0.11]\right).$$

This model is a slightly modified version of the one used in [1]. In all simulations, the number of iterations for the EM algorithm is fixed at $J = 100$. We used the α values of $\alpha = 0 : 0.1 : 1$ for the interval estimates. Figure 1 shows a sample of realizations of the minimum, true and maximum observed output data y_t respectively, while Figure 2 shows the corresponding interval EM estimate of the output observation. Figure 3 compares the maximum and estimated observed output signals, and Figure 4 shows the minimum and estimated observed output signal. Furthermore, we computed the mean square error
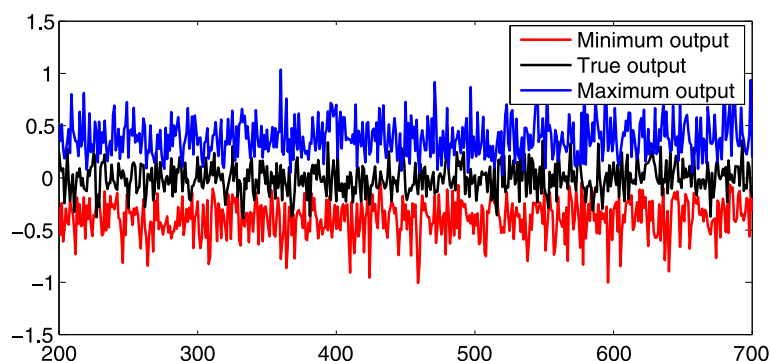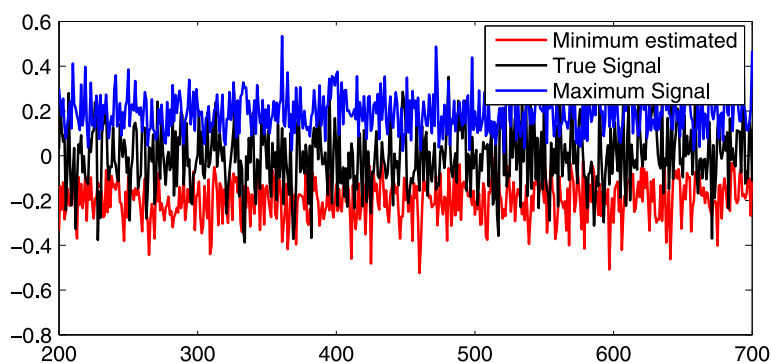
Figure 1 Output signals.
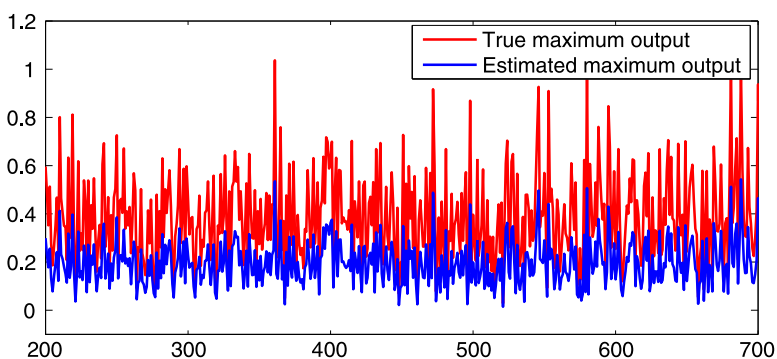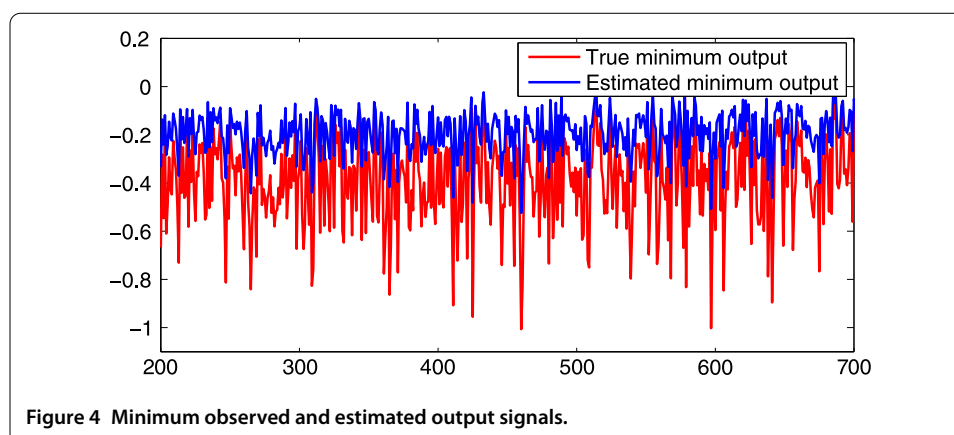

Figure 2 Estimated output signals.

Figure 3 Maximum observed and estimated output signals.

(MSE):

$$E_N = \frac{1}{N} \sum_{t=1}^{N} (y_t - Cx_{t|t-1})^2$$

between the maximum observed and estimated output. The 500 run gave an MSE of 0.0445. A similar computation for the minimum gave an MSE of 0.0447.

Figure 4 Minimum observed and estimated output signals.

Authors' contributions

The authors have achieved equal contributions to each part of this paper. All authors read and approved the final version of the manuscript.

Author details

[1]Department of Mathematics, King Saud University, Riyadh, Kingdom of Saudi Arabia. [2]Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhaharan, 31261, Kingdom of Saudi Arabia. [3]Department of Mathematics, Texas A&M University-Kingsville, Kingsville, TX 78363-8202, USA.

References

1. Chen, G, Wang, J, Shieh, LS: Interval Kalman filtering. IEEE Trans. Aerosp. Electron. Syst. **33**(1), 250-259 (1997)
2. Alefeld, G, Herzberger, J: Introduction to Interval Computations. Academic Press, San Diego (1983)
3. Rohn, J: Inverse interval matrix. SIAM J. Numer. Anal. **3**, 864-870 (1993)
4. Bentbib, AH: Conjugate directions method for solving interval linear systems. Numer. Algorithms **21**, 79-86 (1999)
5. Kubica, BJ, Malinowski, K: Interval random variables and their application in queueing systems with long-tailed service times. In: SMPS, pp. 393-403 (2006)
6. Chen, W, Tan, S: Robust portfolio selection using interval random programming. In: FUZZ-IEEE, Korea, 2009, August 20-24 (2009)
7. Aubin, J-P, Frankowska, H: Set-Valued Analysis. Birkhäuser, Basel (1990)
8. Ekland, I, Témam, R: Convex Analysis and Variational Problems. Classics in Applied Mathematics, vol. 28. SIAM, Philadelphia (1999)
9. Jazwinski, A: Stochastic Precesses and Filtering Theory. Academic Press, New York (1970)
10. Tanizaki, H: Nonlinear Filtering: Estimation and Applications. Springer, Berlin (1996)
11. Bilms, JA: A gentle tutorial of the EM algorithm and its applications to parameter estimation for Gaussian mixture and hidden Markov models. Technical report TR-97-021, ICSI (1997)
12. Dempster, AP, Laird, NM, Rubin, DB: Maximum likelihood from uncomplete data via the EM algorithm. J. R. Stat. Soc. B **39**, 1-38 (1977)
13. Shumway, RH, Stoffer, DS: An approach to time series smoothing and forecasting using the EM algorithm. J. Time Ser. Anal. **3**(4), 253-264 (1982)
14. Shumway, RH, Stoffer, DS: Time Series Analysis and Its Applications. Springer, Berlin (2006)