

# Filtering with clouds

Sebastien Destercke · Olivier Strauss

Published online: 2 November 2011  
© Springer-Verlag 2011

**Abstract** Selecting a particular kernel to filter a given digital signal can be a difficult task. One solution to solve this difficulty is to filter with multiple kernels. However, this solution can be computationally costly. Using the fact that most kernels used for low-pass signal filtering can be assimilated to probability distributions (or linear combinations of probability distributions), we propose to model sets of kernels by convex sets of probabilities. In particular, we use specific representations that allow us to perform a robustness analysis without added computational costs. The result of this analysis is an interval-valued filtered signal. Among such representations are possibility distributions, from which have been defined maxitive kernels. However, one drawback of maxitive kernels is their limited expressiveness. In this paper, we extend this approach by considering another representation of convex sets of probabilities, namely clouds, from which we define cloudy kernels. We show that cloudy kernels are able to represent sets of kernels whose bandwidth is upper and lower bounded, and can therefore be used as a good trade-off between the classical and the maxitive approach, avoiding some of their respective shortcomings without making

computations prohibitive. Finally, the benefits of using cloudy filters is demonstrated through some experiments.

**Keywords** Signal processing · Interval-valued fuzzy sets · Generalised p-boxes

## 1 Introduction

Reconstructing a continuous signal from a set of sampled and possibly corrupted observations is a common problem in both digital analysis and signal processing (Jan 2000). In this context, kernel-based methods can be used for different purposes: reconstruction, impulse response modelling, interpolation, linear and non-linear transformations, stochastic or band-pass filtering, etc.

Most kernels used in signal processing are summative kernels, or a linear combination of summative kernels. A summative kernel is a positive function whose integral is equal to one. A summative kernel is therefore formally equivalent to a probability distribution, and can be identified with it. In practice, summative kernels used for signal filtering are often bounded, continuous, monomodal and symmetric, and we will therefore focus our study on such kernels.

However, how to choose the right kernel together with its parameters to filter a given signal is often a tricky question. To overcome this difficulty, Loquin and Strauss (2008) have proposed to use maxitive kernels instead of summative kernels. Maxitive kernels are formally equivalent to possibility distributions (Dubois and Prade 1988), that can be used to model convex sets of summative kernels (or sets of probability distributions; Dubois and Prade 1992) having an upper-bounded bandwidth. Maxitive kernels can be used to perform a robustness or sensitivity

---

This paper is a revised and extended version of Destercke and Strauss (2010).

---

S. Destercke (✉)  
INRA/CIRAD, UMR1208, 2 place P. Viala,  
34060 Montpellier Cedex 1, France  
e-mail: sebastien.destercke@cirad.fr

O. Strauss  
Laboratoire d'Informatique,  
de Robotique et de Microélectronique de Montpellier (LIRMM),  
161 rue Ada, Montpellier Cedex 5, France  
e-mail: olivier.strauss@lirmm.fr

analysis of the filtering process in a computationally efficient way. This analysis has some interesting features: for example, the maxitive kernel-based filtering approach comes with quantification of the noise level altering the considered signal (Loquin and Strauss 2009).

The output signal obtained using a maxitive kernel on an input digital signal is interval-valued. The bounds of this signal correspond to the envelope of output signals that would have been obtained by filtering with the corresponding set of summative kernels. In this sense, the maxitive kernel approach and its generalisation presented in this paper are different from fuzzy filtering approaches (Aja-Fernandez et al. 2003; Toprak and Guler 2007) or usual robust filtering approaches (Grigoryan and Dougherty 2001; Ma et al. 2010). Indeed, these approaches aim at obtaining an optimal and flexible filter able to cope with situations deviating from the model initial hypothesis, but still deliver a point-valued signal as their output.

Two of the main interests of using maxitive kernels as robust models are their simplicity of representation and the low computational cost associated with the estimation of the (interval-valued) filtered signal. The price to pay for such features is a limited expressiveness, i.e., the fact that it may be impossible to exclude some unwanted summative kernels from the represented set of summative kernels. For instance, as the bandwidth of summative kernels included in maxitive kernels is not lower bounded, this set always includes the Dirac measure.

To overcome this shortcoming of maxitive kernels while keeping their interesting features, we propose to use another uncertainty representation called clouds (Neumaier 2004), calling the kernels defined from them *cloudy kernels*. Cloudy kernels can act as a compromise between summative and maxitive kernels. As we shall see, the interest of using cloudy kernels is twofold: first, they are more expressive than maxitive kernels, the latter being a special case of the former (Destercke et al. 2008), and can take account of additional information or wanted features; second, the computational complexity associated with their use remains very low, an important feature in signal processing.

Nowadays signal processing is usually achieved using computers: the signal to be processed is a digital signal, i.e. a sampled and quantised version of the real continuous signal. We thus consider, in this paper, an algorithmic approach equivalent to the usual signal processing methods that go from a continuous to a discrete setting (see Unser et al. 1993, for example).

Using sets of kernels within a discrete setting can also model an imperfectly known sampling process. Indeed, perfect sampling is usually modelled by the multiplication of the continuous signal with a bounded Dirac comb; however, such an idealistic situation barely exists: the measurement devices and the analog to digital converters

generally induce a smoothing effect. In theory, this smoothing effect can be easily modelled by convoluting the signal with the impulse response of the sensory device (measurement and converter), this impulse response being more often than not a summative kernel. Thus, going from continuous to discrete convolution involves convoluting both the involved filtering kernel and the signal with a smoothing (summative) kernel, provided this latter one is known (Bracewell 1965). However, when considering a digital signal, the sampling kernel which is the impulse response of the sensory device is often unknown. Then digitalising the kernel should rather involve a whole family of possible sampling kernels.

We start by introducing summative and maxitive kernels, before showing how cloudy kernels can act as intermediate representations between them (Sect. 2). Section 3 then studies the computational aspects of using cloudy kernels and provides an efficient algorithm to perform signal filtering with such kernels. The results of some experiments on different signals are then discussed (Sect. 4).

## 2 Between summative and maxitive kernels: cloudy kernels

In this section, we review the basics about summative and maxitive kernels, before introducing cloudy kernels (i.e., kernels based on clouds). We then relate them to the two former representations, and show that they can be used to model sets of summative kernels with a lower-bounded (and upper-bounded) bandwidth.

For readability purposes, we will restrict ourselves to representations on the real line  $\mathbb{R}$  and its discretisation  $\mathcal{X}$ . However, extensions of the presented methods to some product space  $\mathbb{R}^p$  is straightforward.

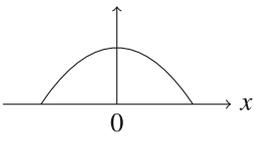
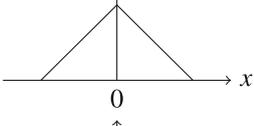
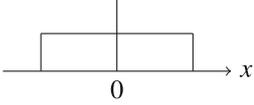
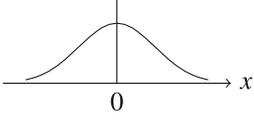
### 2.1 Summative kernels

We define a summative kernel  $\mu$  as a Lebesgue-measurable positive function  $\mu : \mathbb{R} \rightarrow \mathbb{R}^+$  satisfying the normalisation condition  $\int_{-\infty}^{\infty} \mu(x) dx = 1$ . It is formally equivalent to a probability distribution on the real line and can be interpreted as such. The associated probability measure, a function  $P_\mu : 2^{\mathbb{R}} \rightarrow [0, 1]$  from the measurable subsets of  $\mathbb{R}$  to the unit interval, is such that for any measurable subset  $A \subseteq \mathbb{R}$  (also called an event) we have

$$P_\mu(A) = \int_A \mu(x) dx.$$

In this paper, we often consider families of bounded, continuous, symmetrical and monomodal kernels parameterised by their bandwidth. In order to improve readability,

**Table 1** Some classical summative kernels

| Name               | $\kappa_\Delta$   | Shape   |
|--------------------|---|---|
| Epanechnikov       | $\kappa_\Delta(x) = \frac{3}{4\Delta} \left(1 - \left(\frac{x}{\Delta}\right)^2\right) I_\Delta$            |  |
| Triangular         | $\kappa_\Delta(x) = \left(1 - \left \frac{x}{\Delta}\right \right) I_\Delta$                                |  |
| Uniform            | $\kappa_\Delta(x) = \frac{1}{2\Delta} I_\Delta$   |  |
| Truncated Gaussian | $\kappa_\Delta(x) = \frac{4}{\Delta\sqrt{2\pi}} \exp\left(-\left(\frac{x}{\Delta}\right)^2\right) I_\Delta$ |  |

we will use the notation  $\kappa$  when referring to kernels belonging to such families, while keeping  $\mu$  as a notation for generic summative kernels. Let  $\kappa$  denote a basic continuous summative kernel such that  $\kappa(x) = \kappa(-x)$ , whose support is  $[-1, 1]$ . We denote by  $\kappa_\Delta$  the summative kernel derived from  $\kappa$  by  $\kappa_\Delta(x) = \frac{1}{\Delta} \kappa\left(\frac{x}{\Delta}\right)$ . This means that the kernel  $\kappa_\Delta$  has a bandwidth  $\Delta$  and is defined on a compact interval  $[-\Delta, \Delta] \subseteq \mathbb{R}$  centred around zero. Typical kernels belonging to such families are recalled and represented in Table 1.

To a summative kernel  $\kappa_\Delta$  can be associated its (continuous) cumulative distribution function  $F_{\kappa_\Delta} : [-\Delta, \Delta] \rightarrow [0, 1]$ . For any  $x \in [-\Delta, \Delta]$ , we have

$$F_{\kappa_\Delta}(x) = \int_{-\Delta}^x \kappa_\Delta(x) dx = P_{\kappa_\Delta}([-\Delta, x]), \tag{1}$$

and  $F_{\kappa_\Delta}$  is such that  $F_{\kappa_\Delta}(0) = 1/2$  and  $F(x) + F(-x) = 1$ . Similarly, to any summative kernel  $\mu$  its cumulative distribution  $F_\mu$  can be associated with Eq. 1.

### 2.2 Maxitive kernels

A maxitive kernel  $\pi$  is a normalised function  $\pi : \mathbb{R} \rightarrow [0, 1]$  with at least one  $x \in \mathbb{R}$  such that  $\pi(x) = 1$ . A maxitive kernel can be associated with a possibility distribution (Dubois and Prade 1988) and its two (lower and upper) confidence measures, respectively, called necessity and possibility measures. These measures are dual (in the sense that providing one of them on all events is sufficient to

retrieve the other measure) and are such that, for any event  $A \subseteq \mathbb{R}$ , we have:

$$\Pi(A) = \sup_{x \in A} \pi(x), \quad N(A) = 1 - \Pi(A^c) = \inf_{x \in A^c} (1 - \pi(x)), \tag{2}$$

with  $A^c$  the complement of  $A$ . The properties of these lower and upper confidence measures is what make maxitive kernels instrumental and computationally tractable tools to filter signals with sets of kernels. Note that a maxitive kernel is formally equivalent to a fuzzy set (Zadeh 1978). From a basic maxitive kernel  $\pi$  whose support is  $[-1, 1]$ , another maxitive kernel  $\pi_\Delta$  whose bandwidth is  $[-\Delta, \Delta]$  can be computed by the following equation:

$$\pi_\Delta(x) = \pi\left(\frac{x}{\Delta}\right).$$

A maxitive kernel defines a convex set of summative kernels  $\mathcal{P}_\pi$  whose associated probability measures are bounded by the necessity and possibility measures induced by  $\pi$ :

$$\mathcal{P}_\pi = \left\{ \mu \in \mathbb{P}_\mathbb{R} \mid \forall A \subseteq \mathbb{R}, \quad N(A) \leq P_\mu(A) \leq \Pi(A) \right\},$$

with  $\mathbb{P}_\mathbb{R}$  being the set of all summative kernels over  $\mathbb{R}$ ,  $\Pi(A)$  and  $N(A)$  being the possibility and necessity measures induced by  $\pi$ . If a given summative kernel  $\mu$  is in  $\mathcal{P}_\pi$ , we say, by a small abuse of language, that  $\pi$  includes  $\mu$  (or that  $\mu$  is included in  $\pi$ ). Moreover, if a kernel  $\kappa$  is included in  $\pi$  then  $\kappa_\Delta$  is included in  $\pi_\Delta$ . This particular interpretation, together with the fact that Eq. 2 is simple to evaluate (compared to the evaluation of a probability measure from a summative kernel, it simply consists in replacing the summation with a maximum), makes maxitive kernels instrumental tools to filter signals when the identification of a single summative kernel is difficult (Loquin and Strauss 2008).

There are many ways to construct a maxitive kernel that includes a given set of summative kernels (Baudrit and Dubois 2006, Dubois et al. 2004). Here, we will consider the so-called Dubois–Prade transformation. This transformation provides a way to build the most specific maxitive kernel  $\pi_\mu$  including a given summative kernel  $\mu$  [in the sense that any maxitive kernel  $\pi'$  such that  $\pi' \leq \pi_\mu$  with at least one  $x$  such that  $\pi'(x) < \pi_\mu(x)$  does not include  $\mu$ ]. When one wants to build a maxitive kernel including a set  $\mathcal{P}$  of summative kernels, it is then sufficient to take the maximum of each maxitive kernel built from each summative kernel in  $\mathcal{P}$ , using the Dubois–Prade transformation each time.

When summative kernels to include in a maxitive kernel belong to a particular family  $\kappa_\Delta$ , we just need to consider the Dubois–Prade transformation of the summative kernel with the largest bandwidth, that is the one that gives the

least specific maxitive kernel. This corresponds to the case where the shape of the suitable summative kernel is known but where the suitable bandwidth is ill-known. It is then sufficient to consider the kernel  $\kappa_\Delta$  having the maximal bandwidth and to apply the Dubois–Prade transformation to obtain a maxitive kernel including all other summative kernels  $\kappa_{\Delta'}$  with  $\Delta' \leq \Delta$ . Also, in this case the Dubois–Prade transformation can be formulated in a simple way. Given a summative kernel  $\kappa_\Delta$ , the maxitive kernel  $\pi_{\kappa_\Delta}$  resulting from the Dubois–Prade transformation is such that

$$\pi_{\kappa_\Delta}(x) = \begin{cases} 2 \times F_{\kappa_\Delta}(x) & \text{if } x \leq 0 \\ 2(1 - F_{\kappa_\Delta}(x)) & \text{if } x > 0 \end{cases}$$

And  $\pi_{\kappa_\Delta}^+, \pi_{\kappa_\Delta}^-$  denote the following functions

$$\pi_{\kappa_\Delta}^-(x) = \begin{cases} \pi_{\kappa_\Delta}(x) & \text{if } x \leq 0 \\ 1 & \text{if } x > 0, \end{cases} \tag{3}$$

$$\pi_{\kappa_\Delta}^+(x) = \begin{cases} 1 & \text{if } x \leq 0 \\ \pi_{\kappa_\Delta}(x) & \text{if } x > 0. \end{cases} \tag{4}$$

The convex set  $\mathcal{P}_{\pi_{\kappa_\Delta}}$  includes, among others, all summative kernels  $\kappa_{\Delta'}$  with  $\Delta' \in [0, \Delta]$  (Baudrit and Dubois 2006). Among such summative kernels is the Dirac distribution centered in 0, denoted by  $\delta_0$ . This means that the use of maxitive kernels allows us to consider families of kernels whose bandwidths are upper-bounded, but not lower-bounded. This is clearly a shortcoming of maxitive kernels, as in many applications involving signal filtering, the use of the Dirac measure is unwanted (e.g., for modelling a set of smoothing kernels). In such cases, it is desirable to consider families of kernels where the bandwidth is both lower- and upper-bounded.

In the next sections, we show that the recent uncertainty representation called clouds can meet this requirement while preserving computational efficiency. We call (imprecise) kernels derived from such representation *cloudy kernels*.

### 2.3 Cloudy kernels

Cloud, the uncertainty representation used to model cloudy kernels, was introduced by Neumaier (2004) as a way to deal with imprecise probabilistic knowledge. Clouds on the real line are defined as follows:

**Definition 1** A cloud is a pair of mappings  $[\pi, \eta]$  from  $\mathbb{R}$  to the unit interval  $[0, 1]$  such that  $\eta \leq \pi$  and there is at least one element  $x \in \mathbb{R}$  such that  $\pi(x) = 1$  and one element  $y \in \mathbb{R}$  such that  $\eta(y) = 0$ .

Following Neumaier (2004), a cloud  $[\pi, \eta]$  induces a probability family  $\mathcal{P}_{[\pi, \eta]}$  such that

$$\mathcal{P}_{[\pi, \eta]} = \left\{ \mu \in \mathbb{P}_{\mathbb{R}} \mid P_\mu(\{x \mid \eta(x) \geq \alpha\}) \leq 1 - \alpha \leq P_\mu(\{x \mid \pi(x) > \alpha\}) \right\} \tag{5}$$

Similarly to the necessity and possibility measures of maxitive kernels,  $\mathcal{P}_{[\pi, \eta]}$  induces lower and upper confidence measures  $\underline{P}_{[\pi, \eta]}, \bar{P}_{[\pi, \eta]}$  such that, for any event  $A \subseteq \mathbb{R}$ ,

$$\underline{P}_{[\pi, \eta]}(A) = \inf_{\mu \in \mathcal{P}_{[\pi, \eta]}} P_\mu(A) \quad \text{and} \quad \bar{P}_{[\pi, \eta]}(A) = \sup_{\mu \in \mathcal{P}_{[\pi, \eta]}} P_\mu(A).$$

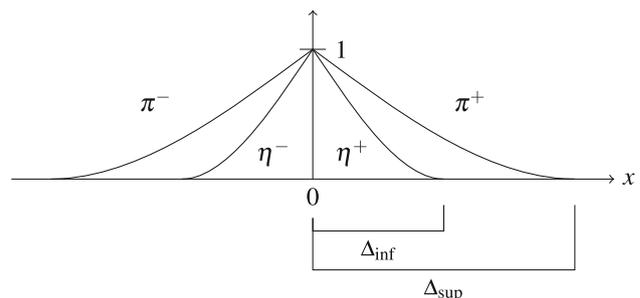
Also note that, formally, clouds are equivalent to interval-valued fuzzy sets (Dubois and Prade 2005) satisfying additional boundary conditions [i.e.,  $\pi(x) = 1$  and  $\eta(y) = 0$ ]. They thus provide a semantic interpretation for this type of modelling in terms of uncertainty representations. A particularly interesting family of clouds are *comonotonic clouds*, also called generalised p-boxes (Destercke et al. 2008) (due to their relationship with another popular uncertainty model called p-box; Ferson et al. 2003). Comonotonic clouds are defined as follows:

**Definition 2** As cloud is said to be comonotonic if  $\forall x, y \in \mathbb{R}, \pi(x) < \pi(y) \Rightarrow \eta(x) \leq \eta(y)$ .

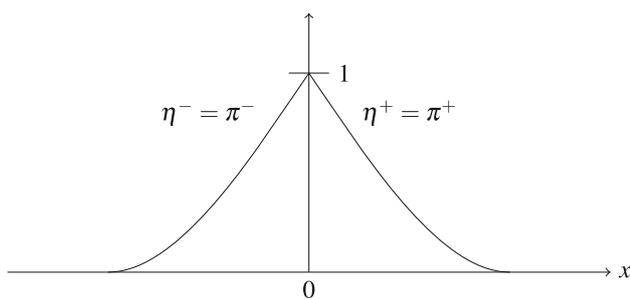
A *cloudy kernel* is simply a pair of functions  $[\pi, \eta]$  that satisfies Definition 1. The notion of cloudy kernel is illustrated in Fig. 1. As for maxitive kernels, we can associate cloudy kernels with sets of summative kernels by identifying  $\mathcal{P}_{[\pi, \eta]}$  with the corresponding set of summative kernels. In this paper, we will restrict ourselves to cloudy kernels represented by continuous, bounded, symmetric and unimodal comonotonic clouds. Again, to make the notations easier, we will consider that they are defined on the interval  $[-\Delta, \Delta]$ .

**Definition 3** A continuous, unimodal and symmetric cloudy kernel defined on  $[-\Delta, \Delta]$  is such that, for any  $x \in [-\Delta, \Delta], \eta(x) = \eta(-x), \pi(x) = \pi(-x)$  and  $\eta, \pi$  are continuous non-decreasing (non-increasing) in  $[-\Delta, 0]([0, \Delta])$ .

Such a cloudy kernel is pictured in Fig. 1. As done in the case of maxitive kernels with Eqs. 3 and 4, given a



**Fig. 1** Example of a cloudy kernel



**Fig. 2** Example of a thin cloud

unimodal symmetric cloudy kernel,  $\eta^+, \eta^-$  denote the functions such that

$$\eta^-(x) = \begin{cases} \eta(x) & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases} \tag{6}$$

$$\eta^+(x) = \begin{cases} 1 & \text{if } x \leq 0 \\ \eta(x) & \text{if } x > 0. \end{cases} \tag{7}$$

Two particular cases of cloudy kernels that will be of interest here correspond to so-called thin and fuzzy clouds.

**Definition 4** A cloudy kernel is said to be thin if  $\forall x \in \mathbb{R}, \pi(x) = \eta(x)$ .

**Definition 5** A cloudy kernel is said to be fuzzy if  $\forall x \in \mathbb{R}, \eta(x) = 0$ .

Figure 2 pictures a thin cloudy kernel. Note that the set of summative kernels  $\mathcal{P}_{[\pi, \eta]}$  modelled by a fuzzy cloudy kernel  $[\pi, \eta]$  coincide with the set modelled by the maxitive kernel  $\pi$  alone (Destercke et al. 2008), showing that maxitive kernels are particular instances of cloudy kernels. We now recall some properties of clouds and cloudy kernels that will be used in this study.

**Proposition 1** (Neumaier 2004) *A cloudy kernel  $[\pi, \eta]$  is included in another one  $[\pi', \eta']$  (in the sense that  $\mathcal{P}_{[\pi, \eta]} \subseteq \mathcal{P}_{[\pi', \eta']}$ ) if and only if, for all  $x \in \mathbb{R}, [\pi(x), \eta(x)] \subseteq [\pi'(x), \eta'(x)]$ .*

Hence, given a cloudy kernel  $[\pi, \eta]$ , any thin cloud  $[\pi', \eta']$  such that  $\eta \leq \eta' = \pi' \leq \pi$  is included in  $[\pi, \eta]$ . Also note that if  $[\pi, \eta]$  is a continuous, symmetric unimodal cloud and  $[\pi', \eta']$  a continuous, symmetric unimodal thin cloud with the same mode as  $[\pi, \eta]$ , then if  $[\pi', \eta]$  does not satisfy this condition [i.e. there is an  $x$  such that  $\eta'(x) < \eta(x)$  or  $\pi'(x) > \pi(x)$ ], we have  $\mathcal{P}_{[\pi, \eta]} \cap \mathcal{P}_{[\pi', \eta]} = \emptyset$ .

**Proposition 2** (Destercke et al. 2008) *The convex set  $\mathcal{P}_{[\pi, \eta]}$  induced by a thin cloudy kernel  $[\pi, \eta]$  includes the two summative kernels having  $F^-, F^+$  for cumulative distributions such that, for all  $x \in \mathbb{R}$*

$$F^-(x) = \eta^-(x) = \pi^-(x), \tag{8}$$

$$F^+(x) = 1 - \eta^+(x) = 1 - \pi^+(x). \tag{9}$$

Note that, since  $\mathcal{P}_{[\pi, \eta]}$  is a convex set, every convex combination of  $F^-, F^+$  is also in the thin cloudy kernel.

### 2.4 Summative kernel approximation with cloudy kernels

Now that cloudy kernels have been introduced, let us show how they can solve the problem occurring with the use of maxitive kernels, i.e., how they can model families of kernels  $\kappa_\Delta$  where  $\Delta$  is lower- and upper-bounded. Note that from now on, we will only deal with continuous, unimodal and symmetric cloudy kernels (kernels satisfying Definition 3) and their discretisation. Assume that we want a model including the family of kernels  $\kappa_\Delta$  such that  $\Delta \in [\Delta_{\text{inf}}, \Delta_{\text{sup}}]$ , and no other kernel of this family with  $\Delta$  outside this interval. To satisfy this requirement, we propose to consider the cloudy kernel  $[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}$  such that, for any  $x \in \mathbb{R}$ :

$$\pi_{\Delta_{\text{sup}}}(x) = \begin{cases} 2 \times F_{\Delta_{\text{sup}}}(x) & \text{if } x \leq 0 \\ 2(1 - F_{\Delta_{\text{sup}}}(x)) & \text{if } x \geq 0 \end{cases} \tag{10}$$

$$\eta_{\Delta_{\text{inf}}}(x) = \begin{cases} 2 \times F_{\Delta_{\text{inf}}}(x) & \text{if } x \leq 0 \\ 2(1 - F_{\Delta_{\text{inf}}}(x)) & \text{if } x \geq 0 \end{cases} \tag{11}$$

We now show that this cloudy kernel contains the two kernels having  $\Delta_{\text{inf}}, \Delta_{\text{sup}}$  for bandwidth, as well as all the kernels of the same family having a bandwidth  $\Delta \in [\Delta_{\text{inf}}, \Delta_{\text{sup}}]$ .

**Proposition 3** *The cloudy kernel  $[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}$  includes the two summative kernels  $\kappa_{\Delta_{\text{inf}}}$  and  $\kappa_{\Delta_{\text{sup}}}$ .*

*Proof* To prove this proposition, we will simply show that the cumulative distribution  $F_{\Delta_{\text{inf}}}$  (resp.,  $F_{\Delta_{\text{sup}}}$ ) of  $\kappa_{\Delta_{\text{inf}}}$  (resp.,  $\kappa_{\Delta_{\text{sup}}}$ ) is included in the cloudy kernel  $[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}$ .

First, from the definition of our cloudy kernel, thin cloudy kernels having  $\pi_{\Delta_{\text{sup}}}$  and  $\eta_{\Delta_{\text{inf}}}$  as distributions are both included in  $[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}$  (Proposition 1).

Let  $F_\pi^-, F_\pi^+$  and  $F_\eta^-, F_\eta^+$  denote the cumulative distributions given by Eqs. 8 and 9, respectively, applied to the thin cloudy kernels  $\pi_{\Delta_{\text{sup}}}$  and  $\eta_{\Delta_{\text{inf}}}$ . By Proposition 2, they are included in the cloudy kernel  $[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}$ , and since  $\mathcal{P}_{[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}}$  is a convex set,  $1/2F_\pi^- + 1/2F_\pi^+$  and  $1/2F_\eta^- + 1/2F_\eta^+$  are also included in the kernel. As these two convex mixtures are equal to  $F_{\Delta_{\text{inf}}}, F_{\Delta_{\text{sup}}}$ , this ends the proof.  $\square$

**Proposition 4** *The cloudy kernel  $[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}$  includes any summative kernel  $\kappa_\Delta$  (derived from  $\kappa$ ) such that  $\Delta \in [\Delta_{\text{inf}}, \Delta_{\text{sup}}]$ .*

*Proof* For a given kernel  $\kappa_\Delta$ ,  $F_\Delta$  denotes its cumulative distribution. We know from Proposition 2 that the thin cloudy kernel  $[\pi, \eta]_{F_\Delta}$  such that

$$\pi_\Delta(x) = \begin{cases} 2 \times F_\Delta(x) & \text{if } x \leq 0 \\ 2(1 - F_\Delta(x)) & \text{if } x \geq 0 \end{cases}$$

includes the cumulative distribution  $F_\Delta$ . Moreover,  $F_{\Delta_{\text{inf}}}(x) \leq F_\Delta(x) \leq F_{\Delta_{\text{sup}}}(x)$  for  $x \leq 0$ , and  $F_{\Delta_{\text{sup}}}(x) \leq F_\Delta(x) \leq F_{\Delta_{\text{inf}}}(x)$  for  $x \geq 0$ , due to the symmetry of considered summative kernels. This means that  $\pi_{\Delta_{\text{sup}}} \leq \pi_\Delta \leq \eta_{\Delta_{\text{inf}}}$ , therefore the thin cloudy kernel  $[\pi, \eta]_{F_\Delta}$  is included in  $[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}$ , and this ends the proof.  $\square$

These two propositions show that cloudy kernels can be built to include all summative kernels that have a bandwidth between  $\Delta_{\text{inf}}$  and  $\Delta_{\text{sup}}$ . However, this property is also fulfilled by maxitive kernels. Let us now show that, in contrast with maxitive kernels, they can be built to exclude summative kernels with a bandwidth smaller than  $\Delta_{\text{inf}}$ , including the Dirac measure.

**Proposition 5** *A kernel  $\kappa_\Delta$ , derived from  $\kappa$ , having a bandwidth  $\Delta$  such that  $\Delta < \Delta_{\text{inf}}$  or  $\Delta > \Delta_{\text{sup}}$  is not included in the cloudy kernel  $[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}$ .*

*Proof* In the case of  $\Delta < \Delta_{\text{inf}}$ , we have  $F_\Delta(x) \leq F_{\Delta_{\text{inf}}}(x)$  for  $x \leq 0$ , and  $F_\Delta(x) \geq F_{\Delta_{\text{sup}}}(x)$  for  $x \geq 0$  (with at least one  $x \in \mathbb{R}$  such that the inequality is strict). When  $\Delta > \Delta_{\text{sup}}$ , we have  $F_\Delta(x) \geq F_{\Delta_{\text{sup}}}(x)$  for  $x \leq 0$ , and  $F_\Delta(x) \leq F_{\Delta_{\text{inf}}}(x)$  for  $x \geq 0$  (with at least one  $x \in \mathbb{R}$  such that the inequality is strict). Hence, the thin cloudy kernels  $[\pi, \eta]_{F_\Delta}$  such that

$$\pi_\Delta(x) = \begin{cases} 2 \times F_\Delta(x) & \text{if } x \leq 0 \\ 2(1 - F_\Delta(x)) & \text{if } x \geq 0 \end{cases}$$

does not satisfy Proposition 1, so  $\kappa_\Delta$  is not included in  $\mathcal{P}_{[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}}$  when  $\Delta < \Delta_{\text{inf}}$  or  $\Delta > \Delta_{\text{sup}}$ .  $\square$

Hence, using cloudy kernels allows us to remove some of the undesired kernels included in maxitive kernels. Still, as for maxitive kernels, other kernels than the summative kernels of the  $\kappa_\Delta$  family are included in  $\mathcal{P}_{[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}}$ . However, the next proposition shows that using cloudy kernels also limits the bandwidth of such summative kernels.

**Proposition 6** *Any summative kernel  $\mu$ , not derived from  $\kappa$  and included in  $\mathcal{P}_{[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}}$  has a bandwidth  $\Delta \in [\Delta_{\text{inf}}/2, \Delta_{\text{sup}}]$*

*Proof* The fact that the bandwidth of any kernel in  $\mathcal{P}_{[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}}$  is bounded above by  $\Delta_{\text{sup}}$  follows from the fact that  $[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}$  is included in the maxitive kernel modelled by  $\pi_{\Delta_{\text{sup}}}$ .

Now, let us prove that the bandwidth of any kernel in  $\mathcal{P}_{[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}}$  is bounded below by  $\Delta_{\text{inf}}/2$ . First, consider inequalities given by Eq. 5 and a level  $\alpha = 1 - \varepsilon$ . Summative kernels  $\mu$  in  $\mathcal{P}_{[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}}$  must satisfy the inequality  $\varepsilon \leq P_\mu([\pi_{\Delta_{\text{sup}}}^-]^{-1}(1 - \varepsilon), (\pi_{\Delta_{\text{sup}}}^+)^{-1}(1 - \varepsilon))$ .

As  $\varepsilon \rightarrow 0$ , the interval  $[(\pi_{\Delta_{\text{sup}}}^-)^{-1}(1 - \varepsilon), (\pi_{\Delta_{\text{sup}}}^+)^{-1}(1 - \varepsilon)]$  tends to  $\{0\}$ , which means that  $P_{[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}}(\{0 \pm \beta\}) > 0$  for any  $\beta > 0$ . This means that any summative kernel in  $\mathcal{P}_{[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}}$  must be strictly positive in the immediate neighborhood of the point  $\{0\}$ .

Now, still consider inequalities given by Eq. 3 and a level  $\alpha = \varepsilon$ . Summative kernels  $\mu$  in  $\mathcal{P}_{[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}}$  must satisfy the inequality

$$\varepsilon \leq P_\mu([\eta_{\Delta_{\text{inf}}}^-]^{-1}(\varepsilon), (\eta_{\Delta_{\text{inf}}}^+)^{-1}(\varepsilon))^c,$$

with  $A^c$  being the complement of  $A$ . As  $\varepsilon \rightarrow 0$ , the interval  $[(\eta_{\Delta_{\text{inf}}}^-)^{-1}(\varepsilon), (\eta_{\Delta_{\text{inf}}}^+)^{-1}(\varepsilon))^c$  tends to  $[-\Delta_{\text{inf}}, \Delta_{\text{inf}}]^c$ , so  $P_{[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}}([-\Delta_{\text{inf}} + \beta, \Delta_{\text{inf}} - \beta]^c) > 0$  for any  $\beta > 0$ . This means that any summative kernel in  $\mathcal{P}_{[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}}$  must be strictly positive for some values in the immediate neighborhood of either  $\{-\Delta_{\text{inf}}\}$  or  $\{\Delta_{\text{inf}}\}$ .

This shows that the support of any  $\mu$  in  $\mathcal{P}_{[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}}$  is lower-bounded by  $\Delta_{\text{inf}}/2$ , as any  $\mu \in \mathcal{P}_{[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}}$  will be strictly positive around  $\{0\}$  and around either  $\{-\Delta_{\text{inf}}\}$  or  $\{\Delta_{\text{inf}}\}$ .  $\square$

Note that the lower-bound  $\Delta_{\text{inf}}/2$  is actually a minimum. Indeed, the summative kernel having for cumulative distribution  $F = \eta_{\Delta_{\text{inf}}}^-$  is in  $\mathcal{P}_{[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}}$  and has a support equal to  $\Delta_{\text{inf}}/2$ .

### 3 Practical computations

In this section, we discuss how digital filtering can actually be achieved with cloudy kernels while maintaining low computational complexity. Note that while filtering kernels used in a particular problem are usually specified in a continuous setting, their use in computations is discrete. Hence in this section, we consider that we are working on a finite domain  $\mathcal{X}$  of  $N$  elements, that corresponds here to a finite sampling of the (continuous) signal.

We first describe how discretisation of the cloudy kernels introduced in the previous section is done, so that we can move from the continuous to the discrete case. We then explain how classical expectation operators (equivalent to filtering with summative kernels) can be extended through the use of Choquet integrals. We first describe the Choquet integral (Choquet 1954; Denneberg 2000) and its links with

expectation operators, summative kernels and maxitive kernels. We then propose an efficient algorithm for computing this Choquet integral for cloudy kernels.

### 3.1 Discretising cloudy kernels

Let  $[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}$  be the cloudy kernel selected to filter the signal  $f, y \in \mathbb{R}$  the value for which we want to reconstruct the signal value and  $\mathcal{X}$  the set of sampled values [i.e., values  $x$  for which the signal value  $f(x)$  is known] that lies within the interval  $[y - \Delta_{\text{sup}}, y + \Delta_{\text{sup}}]$ .

The discrete cloud  $[\pi, \eta]$  necessary to achieve the computations is built from  $\mathcal{X}$  and  $[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}$  in the following steps:

- let  $x^* \in \mathcal{X} = \arg \max_{x \in \mathcal{X}} \pi_{\Delta_{\text{sup}}}(x)$ . Set  $\pi(x^*) = 1$ ;
- for all  $x \neq x^*$  in  $\mathcal{X}$ , set  $\pi(x) = \pi_{\Delta_{\text{sup}}}(y - x)$ ;
- let  $x_* \in \mathcal{X} = \arg \min_{x \in \mathcal{X}} \eta_{\Delta_{\text{inf}}}(x)$ . Set  $\eta(x_*) = 0$ ;
- for all  $x \neq x_*$  in  $\mathcal{X}$ , set  $\eta(x) = \eta_{\Delta_{\text{inf}}}(y - x)$ .

Values<sup>1</sup>  $\pi(x^*)$  and  $\eta(x_*)$  ensure that the discretised cloudy kernel satisfies Definition 1.

Note that it is possible to define a more conservative discretisation  $[\pi', \eta']$ , i.e. for any  $x \neq x^*$ , define  $\pi'(x) := \min\{\pi(y) | y \in \mathcal{X}, \pi(y) > \pi(x)\}$  and for any  $x \neq x_*$ , define  $\eta'(x) := \max\{\eta(y) | y \in \mathcal{X}, \eta(y) < \eta(x)\}$  ( $\pi'$  and  $\eta'$  being equal to  $\pi$  and  $\eta$  for elements  $x^*$  and  $x_*$ , respectively). If  $N$  is low or if the cloudy kernel  $[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}$  is thin, using this latter discretisation is better as it is a guaranteed outer-approximation of  $[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}$ . However, in practical applications, the cloudy kernel is not thin (otherwise it would be better and simpler to filter with the summative kernel it approximates) and  $N$  is usually sufficiently high, so that the difference between filtering with  $[\pi, \eta]$  or  $[\pi', \eta']$  is negligible. This is why we prefer to discretise  $[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}$  into  $[\pi, \eta]$ , which is computationally less complex to evaluate. Finally, it should be noticed that if  $[\pi, \eta]_{[\Delta_{\text{inf}}, \Delta_{\text{sup}}]}$  is comonotonic, so are  $[\pi, \eta]$  and  $[\pi', \eta']$ .

### 3.2 Expectation operator and Choquet integral

Consider now an arbitrary indexing  $\mathcal{X} = \{x_1, \dots, x_N\}$  of domain  $\mathcal{X}$  elements (not necessarily the usual ordering between real numbers) and a real-valued function  $f$  (here, the sampled values of the signal) on  $\mathcal{X}$ , together with a discretised summative kernel  $\mu_i, i = 1, \dots, N$ , where  $\mu_i = \mu(x_i)$ .

<sup>1</sup> If  $\mathcal{X}$  have multiple elements corresponding to  $\arg \max_{x \in \mathcal{X}} \pi_{\Delta_{\text{sup}}}(x)$  or  $\arg \min_{x \in \mathcal{X}} \eta_{\Delta_{\text{inf}}}(x)$ , for all of them  $\pi(x^*) = 1$  and  $\eta(x_*) = 0$ , respectively.

Classical convolution between the discretised kernel  $\mu$  and the sampled signal  $f$  is equivalent to applying an expectation operator, i.e. computing  $\mathbb{E}_\mu(f)$  such that

$$\mathbb{E}_\mu(f) = \sum_{i=1}^N \mu_i f(x_i).$$

When working with a set  $\mathcal{P}$  of kernels defined on  $\mathcal{X}$ , the expectation operator  $\mathbb{E}$  becomes imprecise, and its result when applied to  $f$  is an interval-valued expectation  $[\underline{\mathbb{E}}(f), \overline{\mathbb{E}}(f)]$  such that

$$\underline{\mathbb{E}}(f) = \inf_{\mu \in \mathcal{P}} \mathbb{E}_\mu(f), \quad \overline{\mathbb{E}}(f) = \sup_{\mu \in \mathcal{P}} \mathbb{E}_\mu(f). \tag{12}$$

In general, these bounds are not easy to compute. However, in some specific cases, practical tools are available that make them easily computable. First recall (Walley 1991) that the lower and upper confidence measures induced by  $\mathcal{P}$  on an event  $A \subseteq \mathcal{X}$  are such that  $\underline{P}(A) = \inf_{\mu \in \mathcal{P}} P_\mu(A)$  and  $\overline{P}(A) = \sup_{\mu \in \mathcal{P}} P_\mu(A)$  and are dual in the sense that  $\underline{P}(A) = 1 - \overline{P}(A^c)$  for any  $A \subseteq \mathcal{X}$ . If  $\underline{P}$  satisfies a property of 2-monotonicity, that is if for any pair  $\{A, B\} \subseteq \mathcal{X}$  we have  $\underline{P}(A \cap B) + \underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B)$ , then Eq. 12 can be solved using the Choquet integral.

Consider a positive bounded function<sup>2</sup>  $f$  on  $\mathcal{X}$ . If  $(\cdot)$  denotes a reordering of elements of  $\mathcal{X}$  such that  $f(x_{(1)}) \leq \dots \leq f(x_{(N)})$ , Choquet integrals giving lower and upper expectations are given by

$$C_{\underline{P}}(f) = \underline{\mathbb{E}}(f) = \sum_{i=1}^N (f(x_{(i)}) - f(x_{(i-1)})) \underline{P}(A_{(i)}),$$

$$C_{\overline{P}}(f) = \overline{\mathbb{E}}(f) = \sum_{i=1}^N (f(x_{(i)}) - f(x_{(i-1)})) \overline{P}(A_{(i)}),$$

with  $f(x_{(0)}) = 0$  and  $A_{(i)} = \{x_{(i)}, \dots, x_{(N)}\}$ . The main difficulty is then to compute the lower and upper confidence measures for the  $N$  sets  $A_{(i)}$ .

### 3.3 Imprecise expectations with cloudy kernels

Since cloudy kernels satisfying Definition 2 (this is the case here) induce lower confidence measures that are  $\infty$ -monotone (Destercke et al. 2008a, b) (an even more restrictive property than 2-monotonicity), the Choquet integral can be used to compute expectations. Let us now see how the lower confidence measures on various events can be efficiently computed (upper confidence measures can be obtained by duality).

<sup>2</sup> Assuming positivity is not constraining here, since if  $c$  is a constant  $\underline{\mathbb{E}}(f + c) = \underline{\mathbb{E}}(f) + c$  and the same holds for  $\overline{\mathbb{E}}$ . Therefore any bounded function can be made positive by a simple translation.

Cloudy kernels  $[\pi, \eta]$  defined on  $\mathcal{X}$  induce a complete pre-order  $\leq_{[\pi, \eta]}$  between elements of  $\mathcal{X}$ , in the sense that  $x \leq_{[\pi, \eta]} y$  if and only if  $\eta(x) \leq \eta(y)$  or  $\pi(x) \leq \pi(y)$ . Given a set  $A \subseteq \mathcal{X}$  and this pre-ordering,  $\underline{x}_A$  and  $\bar{x}_A$  denote, respectively, its lowest and highest elements with respect to  $\leq_{[\pi, \eta]}$ . We now introduce the concepts of  $[\pi, \eta]$ -connected sets, since these sets are instrumental in the computation of confidence measures induced by cloudy kernels.

**Definition 6** Given a cloudy kernel  $[\pi, \eta]$  over  $\mathcal{X}$ , a subset  $C \subseteq \mathcal{X}$  is called  $[\pi, \eta]$ -connected if it contains all elements between  $\underline{x}_C$  and  $\bar{x}_C$ , that is

$$C = \{x \in \mathcal{X} \mid \underline{x}_C \leq_{[\pi, \eta]} x \leq_{[\pi, \eta]} \bar{x}_C\}.$$

Let  $\mathcal{C}$  be the set of all  $[\pi, \eta]$ -connected sets of  $\mathcal{X}$ . Now, any event  $A$  can be inner approximated by another event  $A_*$  such that  $A_* = \bigcup_{C \in \mathcal{C}, C \subseteq A} C$  is the union of all maximal  $[\pi, \eta]$ -connected sets included in  $A$ . Due to an additivity property of the lower confidence measure induced by comonotonic clouds on  $[\pi, \eta]$ -connected sets (Destercke and Dubois 2009), we have

$$\underline{P}(A) = \underline{P}(A_*) = \sum_{C \in \mathcal{C}, C \subseteq A} \underline{P}(C). \tag{13}$$

To simplify the notations used in the filtering algorithm, we consider that elements of  $\mathcal{X}$  are indexed accordingly to  $\leq_{[\pi, \eta]}$ , i.e. elements  $x_1, \dots, x_N$  are indexed from the start such that  $i \leq j$  if and only if  $\eta(x_i) \leq \eta(x_j)$  or  $\pi(x_i) \leq \pi(x_j)$ . Given this ordering, the lower confidence measure of a  $[\pi, \eta]$ -connected set  $C = \{x_i, \dots, x_j\}$  is given by the simple formula<sup>3</sup>

$$\underline{P}(C) = \max\{0, \eta(x_{j+1}) - \pi(x_{i-1})\},$$

with  $\eta(x_{N+1}) = 1$  and  $\pi(x_0) = 0$ . As  $\leq_{[\pi, \eta]}$  is a pre-order, we have to be cautious about possible equalities between some elements. In our case (discretisation of unimodal, symmetric cloudy kernels), at most two elements can be equal with respect to  $\leq_{[\pi, \eta]}$  (this will often be the case, as sampling is often performed at regular time intervals). Figure 3 illustrates a cloudy kernel with 7 (irregularly) sampled values, along with the associated indexing and pre-order.

Algorithm 1 describes how to compute lower confidence measures and the incremental summation giving the lower expectation, while Example 1 provides an illustration of the process. At each step, the  $[\pi, \eta]$ -connected sets forming  $A_{(i)}$  are extracted and the corresponding lower confidence measure is computed. The Choquet integral value is then incremented. Note that two orderings and set of indices are used in the algorithm: the one where elements are ordered

<sup>3</sup> Note that every element  $x \in \mathcal{X}$  such that  $x =_{[\pi, \eta]} x_i$  and  $y \in \mathcal{X}$  such that  $y =_{[\pi, \eta]} x_i$  must be in the connected set.

by values of  $f$ , denoted by  $(\ )$ , and the other where elements are ordered using  $\leq_{[\pi, \eta]}$ , without parenthesis. Unless the function  $f$  is increasingly monotonic in  $\mathbb{R}$ , the indexing following the natural order of numbers is never used.

**Algorithm 1:** Algorithm for lower expectations: basic ideas

```

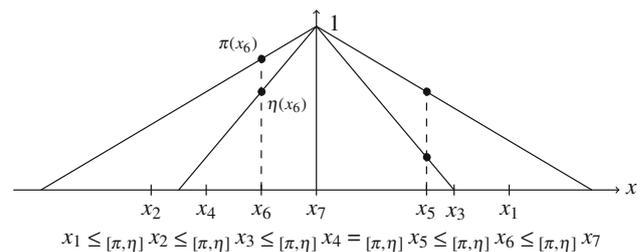
Input:  $f, [\pi, \eta], N$  (number of discretised points)
Output: Lower/upper expectations
 $\underline{\mathbb{E}} = 0$ ;
for  $i = 1, \dots, N$  do
  Compute  $f(x_{(i)}) - f(x_{(i-1)})$ ;
  Extract  $[\pi, \eta]$ -connected sets such that  $A_{(i)} = C_1 \cup \dots \cup C_{M_i}$ ;
  With  $C_j = \{x_k \mid j \leq k \leq \bar{j}\}$ ;
  Compute  $\underline{P}(A_{(i)}) = \sum_{j=1}^{M_i} \max(0, \eta(x_{\bar{j}+1}) - \pi(x_{j-1}))$ ;
 $\underline{\mathbb{E}} = \underline{\mathbb{E}} + [f(x_{(i)}) - f(x_{(i-1)})] \times \underline{P}(A_{(i)})$ 

```

Consider the situation pictured in Fig. 3 (i.e. triangular distributions induced by uniform kernels). For each sampled item  $x$  and associated sampled value  $f(x)$ , we consider that the corresponding distribution values are  $[\eta(x), \pi(x)]$ . Also note that, for symmetrical unimodal cloudy kernels, the values are naturally ordered w.r.t. their distance from the center of the cloud (i.e., the value for which the signal value has to be reconstructed). The values of discretised  $f, \eta$  and  $\pi$  are summarised in Table 2.

If we now apply Algorithm 1 with the lower probability, we get the following steps:

1.  $i = 1, f(x_{(i)}) - f(x_{(i-1)}) = 2, A_{(i)} = \mathcal{X}, \underline{P}(A_{(i)}) = 1 \rightarrow \underline{\mathbb{E}} = 2$
2.  $i = 2, f(x_{(i)}) - f(x_{(i-1)}) = 3, A_{(i)} = C_1 = \{x_2, \dots, x_7\}$   
 $\underline{P}(A_{(i)}) = \max\{0, \eta(x_8) - \pi(x_1)\} = 0.7 \rightarrow \underline{\mathbb{E}} = 3.8$



**Fig. 3** Discretization of cloudy kernels and indexing of elements around  $x_7$  (each  $x_i$  corresponds to a sampled value)

**Table 2** Values for example 1

| Example 1 | $x_1$     | $x_2$     | $x_3$     | $x_4$     | $x_5$     | $x_6$     | $x_7$     |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $\eta$    | 0         | 0         | 0         | 0.2       | 0.2       | 0.6       | 1         |
| $\pi$     | 0.3       | 0.4       | 0.5       | 0.6       | 0.6       | 0.8       | 1         |
| $f$       | 2         | 9         | 8         | 10        | 5         | 6         | 12        |
| $(\ )$    | $x_{(1)}$ | $x_{(5)}$ | $x_{(4)}$ | $x_{(6)}$ | $x_{(2)}$ | $x_{(3)}$ | $x_{(7)}$ |

3.  $i = 3, f(x_{(i)}) - f(x_{(i-1)}) = 1, A_{(i)} = C_1 \cup C_2 = \{x_2, x_3\} \cup \{x_6, x_7\}, \underline{P}(A_{(i)}) = \max\{0, \eta(x_4) - \pi(x_1)\} + \max\{0, \eta(x_8) - \pi(x_5)\} = 0.4 \rightarrow \underline{\mathbb{E}} = 4.2$
4.  $i = 4, f(x_{(i)}) - f(x_{(i-1)}) = 2, A_{(i)} = C_1 \cup C_2 = \{x_2, x_3\} \cup \{x_7\}, \underline{P}(A_{(i)}) = \max\{0, \eta(x_4) - \pi(x_1)\} + \max\{0, \eta(x_8) - \pi(x_6)\} = 0.2 \rightarrow \underline{\mathbb{E}} = 4.2 + 2 \times 0.2 = 4.6$
5.  $i = 5, f(x_{(i)}) - f(x_{(i-1)}) = 1, A_{(i)} = C_1 \cup C_2 = \{x_2\} \cup \{x_7\}, \underline{P}(A_{(i)}) = \max\{0, \eta(x_3) - \pi(x_1)\} + \max\{0, \eta(x_8) - \pi(x_6)\} = 0.2 \rightarrow \underline{\mathbb{E}} = 4.8$
6.  $i = 6, f(x_{(i)}) - f(x_{(i-1)}) = 1, A_{(i)} = C_1 = \{x_7\}, \underline{P}(A_{(i)}) = \max\{0, \eta(x_8) - \pi(x_6)\} = 0.2 \rightarrow \underline{\mathbb{E}} = 5$
7.  $i = 7, f(x_{(i)}) - f(x_{(i-1)}) = 2, A_{(i)} = \{x_7\}, \underline{P}(A_{(i)}) = \max\{0, \eta(x_8) - \pi(x_6)\} = 0.2, \underline{\mathbb{E}} = 5.4$

The lower expectation is finally 5.4. Note that, from step  $i = 3$  to step  $i = 7$ , element  $x_4$  is ignored, due to the fact that  $x_4 = [\pi, \eta] x_5$  (hence, any set including  $x_4$  but not  $x_5$  is treated as if  $x_4$  was not included in it).

#### 4 Experiment: comparison with summative and maxitive kernels

In this section, we illustrate the advantage of using cloudy kernels rather than simple maxitive kernels when filtering a noisy signal. Figure 4 shows a (noisy) signal that has to be filtered by a smoothing kernel. Imprecise kernels (cloudy or maxitive) can be used if the exact shape of the impulse response of the filter is unknown, but it is assumed that this filter is symmetric, centred and has lower and upper bounded bandwidths  $\Delta \in [\Delta_{\text{inf}}, \Delta_{\text{sup}}]$ . Such information can be modelled by a single imprecise kernel and filtering can be achieved by an efficient algorithms, instead of considering multiple filtering with different summative kernels.

The signal pictured in Fig. 4 was obtained by superposing nine sine waves whose frequencies were randomly

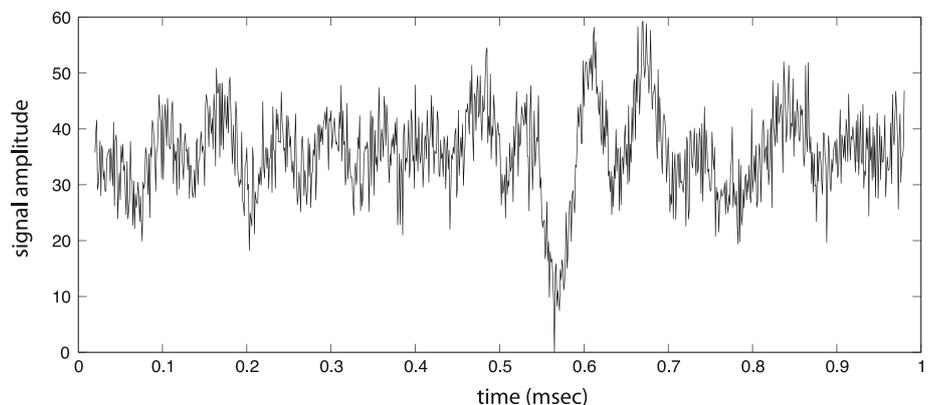
chosen, and were corrupted by adding a normal centred noise with a standard deviation of 5.

We consider here the family of uniform summative kernels with a bandwidth  $\Delta \in [0.018, 0.020]$ . The most specific (triangular) maxitive kernel that dominates this family is the triangular kernel with a bandwidth equal to 0.02, i.e. the maxitive kernel with a bandwidth equating the upper bound of  $\Delta$  (see Loquin and Strauss 2008). The bounds obtained using such a kernel are displayed in Fig. 5 (solid lower and upper lines). As expected, the absence of lower bounds and the inclusion of the Dirac measure inside the maxitive kernel gives very large upper and lower filtered bounds, that encompass the whole signal (i.e. the signal is always in the interval provided by the maxitive kernel). Given our knowledge about the bandwidth, it is clearly desirable to also take account of the lower bound 0.018.

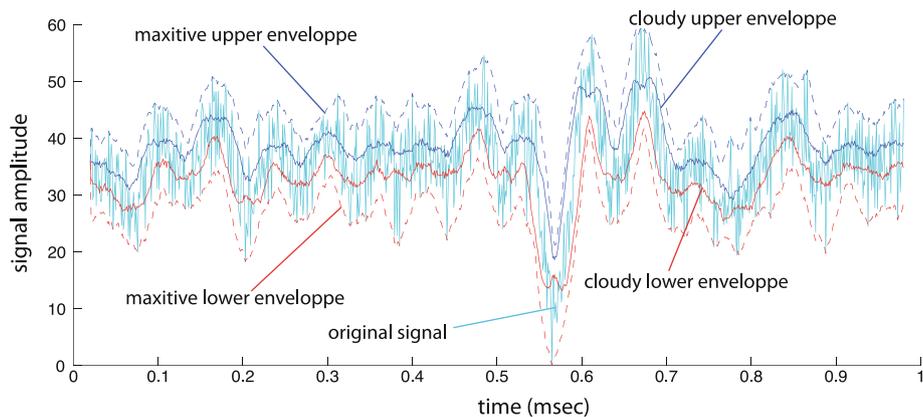
We can fulfil this need using the cloudy kernel presented in this paper. Indeed, a more specific family of kernels that takes the lower bound into account can be obtained using the cloudy kernel composed of two triangular maxitive kernels, with the lower kernel having a bandwidth  $\Delta_{\text{inf}} = 0.018$  and the upper kernel having a bandwidth  $\Delta_{\text{sup}} = 0.020$ . The result of filtering the signal with Algorithm 1 is also pictured in Fig. 5 (dotted upper and lower lines), where we can see that the lower and upper bounds are now much tighter, as expected. Hence, we now have bounds with good confidence levels (as all desired kernels are considered), and which are more informative. Of course, the bounds obtained by cloudy filtering are always included in those obtained by maxitive filtering.

To illustrate the capacity of maxitive and cloudy kernels to encompass the desired kernels, in Fig. 6, we have plotted ten filtered signals obtained using different symmetric centered summative kernels whose bandwidth belongs to the interval  $[\Delta_{\text{inf}}, \Delta_{\text{sup}}]$ . Every filtered signal belongs to the interval-valued signal obtained using the cloudy kernel-based approach.

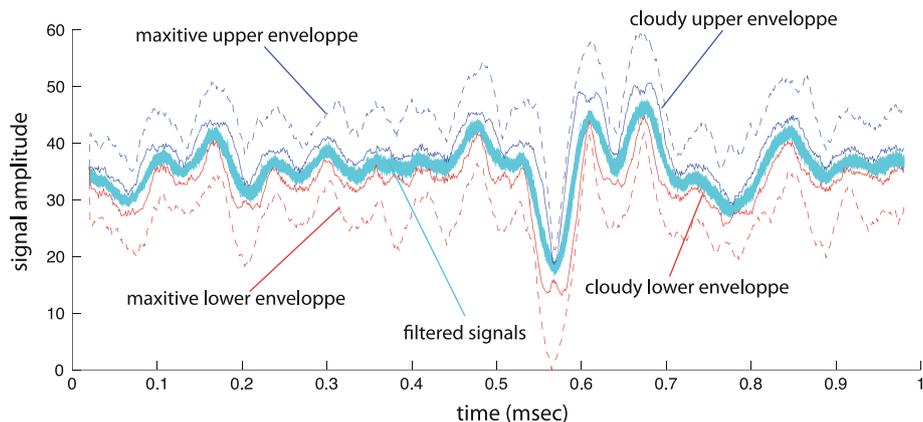
**Fig. 4** Original signal to be filtered



**Fig. 5** Superposition of the original signal, the maxitive imprecise filtering (*dotted upper and lower*) and the cloud-based imprecise filtering (*full upper and lower*)



**Fig. 6** Superposition of nine filtered signals, the maxitive imprecise filtering (*dotted upper and lower*) and the cloud-based imprecise filtering (*full upper and lower*)



## 5 Conclusion

Both summative kernels and sets of summative kernels represented by maxitive kernels have some shortcomings. The former requires choosing a single kernel and bandwidth, which is generally not easy, while the latter often includes unwanted kernels such as the Dirac measure.

In this paper, we have proposed to use cloudy kernels (using the uncertainty representations called clouds) to achieve imprecise linear filtering. Cloudy kernels represent a good trade-off between summative and maxitive kernels, as they allow us to avoid having to choose a single summative kernel while being more expressive than maxitive kernels. This is due to the ability of cloudy kernels to model sets of summative kernels whose bandwidth is both lower- and upper-bounded, while maxitive kernels can only consider upper-bounded bandwidth. We have also proposed simple and efficient (but not necessarily the most efficient) algorithms to compute lower and upper expectations related to cloudy kernels, while keeping a low computational burden on the task of linear filtering.

Our experiments show that cloudy kernels have the expected properties. Compared to summative and maxitive kernels, they allow us to retrieve reliable and informative

envelopes for the filtered signal. However, it appears that envelopes resulting from the use of cloudy kernels are still not very smooth. We suspect that this is due to summative kernels inside the cloudy kernels for which the probability masses are concentrated around some specific points (i.e. mixtures of Dirac measures). To avoid this, we could consider existing techniques (Kozine and Krymsky 2007) to limit the accumulation of such probability masses.

A lot of work is left for future studies, as in the present paper we have only considered families of unimodal bounded centered positive kernels. Many other families of kernels, including kernels having positive and negative values and causal kernels (which are barely symmetric), could be approximated by clouds. However, how to build comontonic clouds that would best represent such families is not straightforward, and could require more ad hoc procedures. In particular, we cannot use a double Dubois–Prade transformation with such families, as they may have many modes that can have different abscissae for different bandwidths.

Another interesting avenue of research would be to combine (or compare) the current approach, which uses imprecise probabilistic representation to model ill-known kernel filters, with the approach proposed by Benavoli

et al. (2009), where imprecise probabilistic models are used to represent noise whose distribution is ill-known.

**Acknowledgments** We wish to thank an anonymous reviewer for his careful reading and insightful comments that have allowed to correct some of the paper contents and improve the presentation.

## References

- Aja-Fernandez S, Alberola-Lopez C, Ruiz-Alzola J (2003) A fuzzy-controlled kalman filter applied to stereo-visual tracking schemes. *Signal Process* 83:101–120
- Baudrit C, Dubois D (2006) Practical representations of incomplete probabilistic knowledge. *Comput Stat Data Anal* 51:86–108
- Benavoli A, Zaffalon M, Miranda E (2009) Reliable hidden Markov model filtering through coherent lower previsions. In: *Proceedings of 12th international conference information fusion*, Seattle, pp 1743–1750
- Bracewell R (1965) *The fourier transform and its applications*. McGraw-Hill, New York
- Choquet G (1954) Theory of capacities. *Annales de l'institut Fourier* 5:131–295
- Denneberg D (2000) Fuzzy measures and integrals—theory and applications, fuzzy measures and integrals—theory and applications. *Physica Verlag, Heidelberg*, pp 42–69
- Destercke S, Dubois D (2009) The role of generalised p-boxes in imprecise probability models. In: Augustin T, Coolen F, Moral S, Troffaes MCM (eds) *Proceedings of the 6th international symposium on imprecise probability: theories and applications*, pp 179–188
- Destercke S, Dubois D, Chojnacki E (2008a) Unifying practical uncertainty representations: I generalized p-boxes. *Int J Approx Reason* 49:649–663
- Destercke S, Dubois D, Chojnacki E (2008b) Unifying practical uncertainty representations: II clouds. *Int J Approx Reason* 49:664–677
- Destercke S, Strauss O (2010) Using cloudy kernels for imprecise linear filtering. In: *13th international conference on information processing and management of uncertainty*. Lecture notes in computer science. Springer, Berlin, pp 198–207
- Dubois D, Foulloy L, Mauris G, Prade H (2004) Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Comput* 10:273–297
- Dubois D, Prade H (1988) *Possibility theory: an approach to computerized processing of uncertainty*. Plenum Press, New York
- Dubois D, Prade H (1992) When upper probabilities are possibility measures. *Fuzzy Sets Syst* 49:65–74
- Dubois D, Prade H (2005) Interval-valued fuzzy sets, possibility theory and imprecise probability. In: *Proceedings of international conference in fuzzy logic and technology (EUSFLAT'05)*, Barcelona, pp 314–319
- Ferson S, Ginzburg L, Kreinovich V, Myers D, Sentz K (2003) *Constructing probability boxes and Dempster-Shafer structures, type*. Technical Report, Sandia National Laboratories
- Grigoryan A, Dougherty E (2001) Bayesian robust optimal linear filters. *Signal Process* 81:2503–2521
- Jan J (2000) *Digital signal filtering, analyses and restoration*. IET, USA
- Kozine I, Krymsky V (2007) Enhancement of natural extension. In: de Cooman G, Vejnarová J, Zaffalon M (eds) *Proceedings of 5th international symposium on imprecise probabilities: theories and applications*, pp 253–262
- Loquin K, Strauss O (2008) On the granularity of summative kernels. *Fuzzy Sets Syst* 159
- Loquin K, Strauss O (2009) Noise quantization via possibilistic filtering. In: *Proceedings of 4th international symposium on imprecise probabilities and their applications*, pp 297–306
- Ma L, Wang Z, Hu J, Bo Y, Guo Z (2010) Robust variance-constrained filtering for a class of nonlinear stochastic systems with missing measurements. *Signal Process* 90
- Neumaier A (2004) Clouds, fuzzy sets and probability intervals. *Reliable Comput* 10:249–272
- Toprak A, Guler I (2007) Impulse noise reduction in medical images with the use of switch mode fuzzy adaptive median filter. *Digital Signal Process* 17:711–723
- Unser M, Aldroubi A, Eden M (1993) B-spline signal processing: part i-theory. *IEEE Trans Signal Process* 41:821–833
- Walley P (1991) *Statistical reasoning with imprecise probabilities*. Chapman and Hall, New York
- Zadeh L (1978) Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst* 1:3–28