# Using atomic bounds to get sub-modular approximations

Sébastien Destercke, Agnès Rico, Olivier Strauss

# Using atomic bounds to get sub-modular approximations

Sébastien Destercke[1], Agnes Rico[2], and Olivier Strauss[3]

[1] UMR CNRS 7253, Heudiasyc, Sorbonne Université, Université de Technologie de Compiègne, France
sebastien.destercke@hds.utc.fr
[2] ERIC, Claude Bernard Université Lyon 1, Lyon, France
Agnes.Rico@univ-lyon1.fr
[3] LIRMM, Univ. Montpellier, CNRS, Montpellier, France
Olivier.Strauss@umontpellier.fr

**Abstract.** Set functions and inferences over them play an important role in many AI-related problems. They are commonly used in kernel-based methods, game theory, uncertainty representation and preference modelling. Of particular importance are additive set-functions and their associated expectation operator, as well as sub-modular functions. However, specifying precisely and completely such set functions can be a daunting task as well as a lot of data and knowledge. When time or information is missing, it may be handy to have approximating tools of such set-functions, that require less information and possibly enjoy nice mathematical properties. In this paper, we show that if the only information we have are atomic bounds of such functions, we can build conservative approximations that are either sub- or super-modular. We then illustrate the potential use of our approximation on a convolution-based signal processing problem.

## 1 Introduction

Set functions or measures play an important role in many aspects of artificial intelligence. The most well-known are undoubtedly probability measures, i.e., additive, normalised positive set functions. Increasing set functions, called capacities or fuzzy measures, generalise such probability measures. They appeared in 50's, and have been extensively used in areas related to decision [1]. In particular, they can represent the weights of coalitions in cooperative games, or measures of uncertainty such as lower probabilities or belief functions (see [12]). Particular fuzzy measures that are maxitive rather than additive are possibility measures [22], the basic building blocks of possibility theory [4,5].

Other capacities that have attracted a lot of attention are sub-modular functions, as they have taken a very important role in many aspects of artificial intelligence. They are used in numerous problems of machine learning and signal processing [18,11], in social choice theory where they can be used to induce fairness in the obtained solutions [3], and in modelling preferences in item selection problems [19]. Indeed, the sub-modularity property comes with many advantages, notably the one to be able to use greedy optimisation algorithms with good guaranteed approximation properties.

There are good reasons to consider set functions that are not necessarily increasing with respect to set inclusion, nor that are normalized as probability or possibility measures can be. A question that follows is whether approximation tools allowing for efficient inferences are still valid in such contexts. This is the question we address in this paper, where we focus on approximation obtained by considering atomic bounds (i.e., lower/upper bounds over singletons).

The paper is structured as follows. Section 2 is a brief recall of set-functions and their use. We will focus on additive set functions and sub-modular set functions. Section 3 is devoted to the approximation of set of additive set-functions and sub-modular functions. Section 4 presents an application to signal convolution.

## 2 Set-functions and their use

In this paper, we are concerned with approximating (sets) of discrete set-functions. By set function $\kappa$ over a finite space $\mathcal{X} = \{x_1, \ldots, x_n\}$, we understand a bounded real-valued function $\kappa : 2^{\mathcal{X}} \to \mathbb{R}$ from the power set of $\mathcal{X}$ to the reals. We furthermore assume that $\kappa(\emptyset) = 0$. The *conjugate* $\kappa^c$ of a set function $\kappa$ is $\kappa^c(A) = \kappa(\mathcal{X}) - \kappa(A^c)$.

### 2.1 Additive set-functions and expectation

One of the most widespread set functions are additive ones, i.e., set functions defined from a mass distribution $\psi : \mathcal{X} \to \mathbb{R}$ assigning a mass to each element, from which we have $\kappa_\psi(A) = \sum_{x \in A} \psi(x)$ (in the rest of the paper, we denote by $\kappa_\psi$ additive measures derived from the mass distribution $\psi$).

Once we are given such a set-function, a very common operator applied to them to make inferences is the linear expectation operator. If we have a real valued bounded function $f : \mathcal{X} \to \mathbb{R}$, such an operator is

$$\mathbb{E}_{\kappa_\psi}(f) = \sum_{x \in \mathcal{X}} f(x)\psi(x) \tag{1}$$

$$= \sum_{i=1}^{n} (f(x_{[i]}) - f(x_{[i-1]}))\kappa_\psi(A_{[i]}), \tag{2}$$

where $[.]$ is the permutation such that $(f(x_{[1]}) \leq \cdots \leq f(x_{[N]}))$, $A_{[i]} = \{x_{[i]}, \ldots, x_{[N]}\}$ with, by convention, $x_{[0]} = 0$ and $A_{[N+1]} = \emptyset$. Equations (2) will be helpful when considering sub-modular functions later on.

If we assume that $\psi$ is non-negative, $\psi \geq 0$, and that $\kappa_\psi(\mathcal{X}) = \sum_{x \in \mathcal{X}} \psi(x) = 1$, we retrieve classical probabilities and (1) is just standard expectations. Also, $\psi$ being positive ensures that the obtained function $\kappa$ is monotonic with respect to inclusion, i.e., $\kappa_\psi(A) \geq \kappa_\psi(B)$ whenever $B \subseteq A$.

In this paper, however, we make neither of these two assumptions, i.e., $\kappa_\psi(\mathcal{X})$ can be any real value, as well as $\psi(x)$ that can be negative. This latter relaxation implies that we may have $B \subseteq A$ but $\kappa_\psi(A) > \kappa_\psi(B)$.

While the latter property may seem odd to some readers, there are good reasons to consider such a relaxation. For instance, deriving a signal through convolution (the application of Section 4 is of this kind), we are considering a non-null function $\psi$ such that

$\kappa(\mathcal{X}) = 0$, meaning that some values of $\psi$ must be negative. Similarly, in social ranking problems [14], the elements $x_i$ can be individuals and one search to assess/order the performances of groups or subsets of individuals. In this case, it can be that two groups $A \subset B$ are such that $B$ performs worse than $A$ ($A \succ B$), for instance because $B$ contains individuals that do not get along together and worsen the whole group performance. Should such a statement be numerically encoded as a set function, we should have $\kappa_\psi(A) > \kappa_\psi(B)$.

*Example 1.* Consider a digital signal $f$ taking values on two successive points:

$$f(x_1) = 3, f(x_2) = 5$$

computing its derivative in $x_2$ comes down to compute $\mathbb{E}_{\kappa_\psi}$ with $\psi(x_1) = -1, \psi(x_2) = 1$, giving $\mathbb{E}_{\kappa_\psi}$=2

### 2.2 Sub-modular set functions

Another kind of set functions that are widely used in artificial intelligence are sub-modular set functions. Sub-modularity means that, for any pair $A, B \subseteq \mathcal{X}$ of event, $\kappa$ is such that

$$\kappa(A \cup B) + \kappa(A \cap B) \leq \kappa(A) + \kappa(B), \tag{3}$$

or equivalently [15, Ch. 44] that, for any $A \subseteq B$ and any $x \notin B$, we have

$$\kappa(A \cup \{x\}) - \kappa(A) \geq \kappa(B \cup \{x\}) - \kappa(B). \tag{4}$$

This latter formulation is sometimes referred as the property of *diminishing return*, modelling the fact that element $\{x\}$ has a higher value when added to a smaller subset.

However, in contrast with the additive case, specifying a sub-modular set-function requires in general to specify or to work with $2^{|\mathcal{X}|}$ values, a daunting quantity. In practice, it may be useful to have approximations requiring to collect and work with a limited number of values. In this paper, we will see that this can be done by simply considering $2|\mathcal{X}|$ values of $\kappa$. Note that the idea of non-monotonic (w.r.t. inclusion), non-additive set functions already appears in [10], precisely to model numerically rankings of subsets.

Sub-modular set functions have many interesting properties, but we will only mention those useful for our purpose, and proved in [7]. A first property is that its *core*

$$\mathcal{C}_\kappa = \{\psi : \forall A \subseteq \mathcal{X}, \kappa_\psi(A) \leq \kappa(A), \psi(\mathcal{X}) = \kappa(\mathcal{X})\}, \tag{5}$$

that is the set of additive measures that is dominated by it is never empty. In particular, this means that sub-modular set functions can be useful to describe or approximate sets of additive set functions, e.g., in convolution problems where the convolution filter is ill-known.

A second property one can also show [7] that applying the Choquet integral to a bounded function $f : \mathcal{X} \to \mathbb{R}$ gives

$$\overline{\mathbb{E}}_\kappa(f) = \sum_{i=1}^{n} (f(x_{[i]}) - f(x_{[i-1]}))\kappa(A_{[i]}), \tag{6}$$

$$= \sup_{\psi \in \mathcal{C}_\kappa} \mathbb{E}_{\kappa_\psi}(f) \tag{7}$$

Similar properties apply to the super-modular $\kappa^c$

$$\underline{\mathbb{E}}_\kappa(f) = \sum_{i=1}^{n} (f(x_{[i]}) - f(x_{[i-1]}))\kappa^c(A_{[i]}), \tag{8}$$

$$= \inf_{\psi \in \mathcal{C}_\kappa} \mathbb{E}_{\kappa_\psi}(f) \tag{9}$$

## 3 Approximating set-functions with atomic bounds

In this section, we provide our main results concerning non-additive set functions, that show that one can build sub-modular approximations by only focusing on atomic bounds of some sets of additive measures or of some initial sub-modular function. This provides a very easy way to get approximations by focusing on only $2n$ values.

### 3.1 Approximating sets of additive set-functions

We first consider the case where we have a set $\mathcal{M}$ of possible additive measures. In signal processing, those may come from the fact that the convoluting kernel has an ill-known bandwidth, while in preference modelling they may represent the doubt an agent has about the strength or rankings of coalitions. $\mathcal{M}$ may also be a neighbourhood around some initial distribution $\psi_0$, used for instance to make a sensitivity/robustness study around $\psi_0$.

However, working with general sets $\mathcal{M}$ is highly inconvenient, whether they are discrete or continuous. We will now show that we can easily obtain a sub-modular approximation of $\mathcal{M}$. To do this, we will denote by

$$\underline{\psi}(x) = \inf_{\psi \in \mathcal{M}} \psi(x),$$
$$\overline{\psi}(x) = \sup_{\psi \in \mathcal{M}} \psi(x).$$

Let us now define the two set functions

$$\underline{\mu}_{\mathcal{M}}(A) = \max\left(\sum_{x \in A} \underline{\psi}(x), \kappa_\psi(\mathcal{X}) - \sum_{x \in A^c} \overline{\psi}(x)\right), \tag{10}$$

$$\overline{\mu}_{\mathcal{M}}(A) = \min\left(\sum_{x \in A} \overline{\psi}(x), \kappa_\psi(\mathcal{X}) - \sum_{x \in A^c} \underline{\psi}(x)\right), \tag{11}$$

with $\kappa_\psi(\mathcal{X})$ the value for any $\psi \in \mathcal{M}$. The subscript $\mu_{\mathcal{M}}$ denotes that $\mu$ originates from a set of additive measures.

*Example 2.* Let us assume that some group of persons want to pack some food before going on a hike. They can choose between ($a$)pples, ($b$)ananas, ($c$)horizo and ($d$)anish pastries, so that $\mathcal{X} = \{a, b, c, d\}$. In order to know what to take (it is impossible to take all four types of food), members of the group are allowed to distribute 10 points between all food types to express their individual preferences. They can also give negative scores (up to -5) if they really dislike a food. We assume that food values are additive. So we know that $\kappa_\psi(\mathcal{X}) = 10$, and the span of scores given by the group to each type of food is summarised in Table 1. Everyone likes apples as even the lower score $\underline{\psi}$ is

| Food ($\mathcal{X}$) | $a$ | $b$ | $c$ | $d$ |
|---|---|---|---|---|
| $\overline{\psi}$ | 5 | 7 | 4 | 4 |
| $\underline{\psi}$ | 1 | -3 | -3 | -2 |

Table 1: Table of scores of Example 2

positive (some more than others), but some dislikes bananas (e.g., because of their taste) while some dislikes chorizo (e.g., as they are vegetarian). We can now try to assess the possible values of various packages to compare them. For instance, evaluating the set $\{a, b, c\}$, we get

$$\overline{\mu}(\{a, b, c\}) = \min(5 + 7 + 4, 10 - (-2)) = 12$$

$$\underline{\mu}(\{a, b, c\}) = \max(1 - 3 - 3, 10 - 4) = 6$$

so that this package is clearly interesting, for instance compared to $\{b, c, d\}$ for which $\overline{\mu}(\{b, c, d\}) = 9$.

We will now study the properties of the set functions given by Equations 10-11. Let us first show that the two set functions are conjugate ones.

**Proposition 1.** *For any $\mathcal{M}$, $\underline{\mu}_\mathcal{M}$ and $\overline{\mu}_\mathcal{M}$ are conjugate.*

*Proof.*

$$\begin{aligned}
\underline{\mu}_\mathcal{M}(A) &= \underline{\mu}_\mathcal{M}(\mathcal{X}) - \underline{\mu}_\mathcal{M}(A^c) \\
&= \kappa_\psi(\mathcal{X}) - \max\{\sum_{x \notin A} \underline{\psi}(x), \kappa_\psi(\mathcal{X}) - \sum_{x \in A} \overline{\psi}(x)\} \\
&= \min\{\kappa_\psi(\mathcal{X}) - \sum_{x \notin A} \underline{\psi}(x), \sum_{x \in A} \overline{\psi}(x)\} = \overline{\mu}_\mathcal{M}(A).
\end{aligned}$$

We can then show that $\overline{\mu}_\mathcal{M}$ can be used as a conservative approximation of $\mu_\mathcal{M}$, by the following proposition.

**Proposition 2.** *We have $\mathcal{M} \subseteq \mathcal{C}_{\overline{\mu}_\mathcal{M}}$.*

*Proof.* The proof simply consists in showing that for any $\psi \in \mathcal{M}$, we have $\psi(A) \leq \overline{\mu}(A)$ for any $A \subseteq \mathcal{X}$. To see this, observe that for any $\psi$, we have $\overline{\psi} \geq \psi$. This means that $\sum_{x\in A} \overline{\psi}(x) \geq \sum_{x\in A} \psi(x)$. Similarly, since $\underline{\psi} \leq \psi$, we have $\sum_{x\notin A} \underline{\psi}(x) \leq \sum_{x\notin A} \psi(x)$, hence $\kappa_\psi(\mathcal{X}) - \sum_{x\in A^c} \underline{\psi}(x) \geq \kappa_\psi(\mathcal{X}) - \sum_{x\in A^c} \psi(x) = \sum_{x\in A} \psi(x)$ where the last equality comes from $\psi$ being additive and $\sum_{x\in\mathcal{X}} \psi(x) = \kappa_\psi(\mathcal{X})$.

We now demonstrate that $\underline{\mu}_{\mathcal{M}}$ is a super-modular set function. Among other things, this ensures that applying Equation (8) (that is computationally cheap to evaluate) to a function $f$ will yield bounds that correspond to an expectation operator applied to an additive measure.

**Proposition 3.** *$\overline{\mu}_{\mathcal{M}}$ is a sub-modular set function.*

*Proof.* Following the ideas of [6] that prove super-modularity of $\underline{\mu}_{\mathcal{M}}$ when $\psi$ are probabilities, we will demonstrate the super-modularity of $\underline{\mu}_{\mathcal{M}}$, which by Proposition 1 is the same as proving sub-modularity of $\overline{\mu}_{\mathcal{M}}$. To do so, we consider arbitrary subsets $A, B \subseteq \mathcal{X}$ and we define the sets $C = A \cap B$ and $D = (A \cup B) \setminus (A \cap B)$. We have $C \cap D = \emptyset$ and $C \cup D = A \cup B$, as well as $\mu_{\mathcal{M}}(A \cap B) + \mu_{\mathcal{M}}(A \cup B) = \mu_{\mathcal{M}}(C) + \mu_{\mathcal{M}}(C \cup D)$. So we are going to prove the property considering $C$ and $D$ subsets such that $C \cap D = \emptyset$.

The idea of the proof is to define an additive set function $\kappa_\psi : 2^{\mathcal{X}} \to \mathbb{R}$ and its mass function $\psi$ such that $\underline{\mu}_{\mathcal{M}}(C) = \kappa_\psi(C)$ and $\underline{\mu}_{\mathcal{M}}(C \cup D) = \kappa_\psi(C \cup D)$ and $\underline{\psi} \leq \psi \leq \overline{\psi}$.

This is sufficient to prove sub-modularity of $\underline{\mu}_{\mathcal{M}}$, since $\kappa_\psi(C) + \kappa_\psi(C \cup D) = \kappa_\psi(B) + \kappa_\psi(A)$ due to its additivity, and $\kappa_\psi(A) \geq \underline{\mu}_{\mathcal{M}}(A)$ and $\kappa_\psi(B) \geq \underline{\mu}_{\mathcal{M}}(B)$, follows from $\underline{\psi} \leq \psi \leq \overline{\psi}$.

There are four possible cases to study (two possible values for $\underline{\mu}_{\mathcal{M}}(C)$ and two for $\underline{\mu}_{\mathcal{M}}(C \cup D)$). All the cases are similar since the sets $C$ and $D$ are such that $C \cap D = \emptyset$. We will therefore provide the details for only one of them.

Let us consider the specific case $\underline{\mu}_{\mathcal{M}}(C) = \sum_{x\in C} \underline{\psi}(x)$ and $\underline{\mu}_{\mathcal{M}}(C \cup D) = \kappa(\mathcal{X}) - \sum_{x\notin C\cup D} \overline{\psi}(x)$. On each $x \in \mathcal{X}$ we are going to define the weights $\psi(x)$ as follows:

- $\forall x \in C$, $\psi(x) = \underline{\psi}(x)$,
- $\forall x \notin C \cup D$, $\psi(x) = \overline{\psi}(x)$,

This ensures that $\kappa_\psi$ coincide with $\underline{\mu}_{\mathcal{M}}(C \cup D)$ and $\underline{\mu}_{\mathcal{M}}(C)$

We are now going to see how we can assign masses $d_x := \psi(x)$ on elements of the subset $D$ so that that $\underline{\psi}(x) \leq \psi(x) = d_x \leq \overline{\psi}(x)$ and $\sum_{x\in\mathcal{X}} = \kappa(\mathcal{X})$. Observe that we still have to assign a total (potentially negative) mass $\kappa(\mathcal{X}) - \sum_{x\notin C\cup D} \overline{\psi}(x) - \sum_{x\in C} \underline{\psi}(x)$ on elements of $D$. To do so, consider first that, given the considered case, we have

$$\underline{\mu}_{\mathcal{M}}(C \cup D) - \underline{\mu}_{\mathcal{M}}(C) = \kappa(\mathcal{X}) - \sum_{x\notin C\cup D} \overline{\psi}(x) - \sum_{x\in C} \underline{\psi}(x).$$

As, $\underline{\mu}_{\mathcal{M}}(C) = \sum_{x \in C} \underline{\psi}(x) \geq \kappa(\mathcal{X}) - \sum_{x \notin C} \overline{\psi}(x)$, we have

$$\underline{\mu}_{\mathcal{M}}(C \cup D) - \underline{\mu}_{\mathcal{M}}(C) \leq$$
$$\kappa(\mathcal{X}) - \sum_{x \notin C \cup D} \overline{\psi}(x) - \kappa(\mathcal{X}) + \sum_{x \notin C} \overline{\psi}(x) \leq \sum_{x \in D} \overline{\psi}(x).$$

Similarly, as $\underline{\mu}_{\mathcal{M}}(C \cup D) = \kappa(\mathcal{X}) - \sum_{x \notin C \cup D} \overline{\psi}(x) \geq \sum_{x \in C \cup D} \underline{\psi}(x)$, we have

$$\underline{\mu}_{\mathcal{M}}(C \cup D) - \underline{\mu}_{\mathcal{M}}(C)$$
$$\geq \sum_{x \in C \cup D} \underline{\psi}(x) - \sum_{x \in C} \underline{\psi}(x) = \sum_{x \in D} \underline{\psi}(x).$$

So $\sum_{x \in D} \underline{\psi}(x) \leq \underline{\mu}_{\mathcal{M}}(C \cup D) - \underline{\mu}_{\mathcal{M}}(C) \leq \sum_{x \in D} \overline{\psi}(x)$. This means that we can divide the difference $\underline{\mu}_{\mathcal{M}}(C \cup D) - \underline{\mu}_{\mathcal{M}}(C)$ (equal to the remaining mass to assign) on $D$ and choose a real $d_x$ for all $x \in D$ such that $\underline{\psi}(x) \leq d_x \leq \overline{\psi}(x)$ and $\sum_{x \in D} d_x = \underline{\mu}_{\mathcal{M}}(C \cup D) - \underline{\mu}_{\mathcal{M}}(C)$.

What Propositions 2 and 3 show is that previous results obtained for the probabilistic case (summative functions $\psi$ summing up to one) readily extend to more general and complex cases, opening up new areas of applications such as the signal treatment one described in Section 4 and already suggested in Example 1.

### 3.2 Approximating sub-modular functions

We will now show that our approximation also works when one starts from a sub-modular function $\kappa$ to approximate. There are at least two ways in which our proposed results can be useful: as an upper bound[4], it can be useful in Branch and Bound techniques applied to sub-modular optimization [20]; as a trace on the singletons of a potentially unknown $\kappa$, it allows to build an approximation of $\kappa$ while collecting only a $2n$ of its values.

We consider a sub-modular $\kappa : 2^{\mathcal{X}} \to \mathbb{R}$. Recall that the conjugate set-function $\kappa^c$ of $\kappa$ is $\forall A, \kappa^c(A) = \kappa(\mathcal{X}) - \kappa(A^c)$. It is easy to check that for any pair $A, B \subseteq \mathcal{X}$, $\kappa^c$ is such that

$$\kappa^c(A \cup B) + \kappa^c(A \cap B) \geq \kappa^c(A) + \kappa^c(B),$$

it is super-modular. $\kappa$ is sub-modular so we have $\kappa^c \leq \kappa$. Now let us assume that we know only $\kappa$ and $\kappa^c$ on each element $x \in \mathcal{X}$. This is equivalent to know $\kappa$ on each $x$ and $\{x\}^c$. In Example 2, this could amount to ask group member the value of every single food element, as well as the value of packages if we remove each food element, assuming that they aggregate values in a sub-modular way. By Equation (4), it means that adding a food element to an existing package is all the less interesting as the package grows bigger, which seems a reasonable assumption.

For all $x \in \mathcal{X}$ let us define[5] $\underline{\psi}^{\kappa}(x) = \kappa^c(\{x\})$ and $\overline{\psi}^{\kappa}(x) = \kappa(\{x\})$. The knowledge of $\underline{\psi}^{\kappa}(x)$ and $\overline{\psi}^{\kappa}(x)$ entails the following property:

[4] or lower bound if we consider the complementary measure

[5] The reason for also calling them $\psi$ is clarified in the proof of Proposition 6

**Proposition 4.** $\sum_{x\in\mathcal{X}} \underline{\psi}^{\kappa}(x) \leq \kappa(\mathcal{X}) \leq \sum_{x\in\mathcal{X}} \overline{\psi}^{\kappa}(x)$.

*Proof.* $\underline{\psi}^{k}(x) \leq \overline{\psi}^{k}(x) = \kappa(\{x\}) = \kappa(\mathcal{X}) - \kappa^{c}(\{x\}^{c}) \leq \kappa(\mathcal{X}) - \sum_{x'\neq x} \underline{\psi}^{k}(x')$, so we have $\sum_{x\in\mathcal{X}} \underline{\psi}^{k}(x) \leq \kappa(\mathcal{X})$.
$\overline{\psi}^{k}(x) \geq \underline{\psi}^{k}(x) = \kappa^{c}(\{x\}) = \kappa(\mathcal{X}) - \kappa(\{x\}^{c}) \geq \kappa(\mathcal{X}) - \sum_{x'\neq x} \overline{\psi}^{k}(x')$, so we have $\kappa(\mathcal{X}) \leq \sum_{x\in\mathcal{X}} \overline{\psi}^{k}(x)$.

As in the previous section, we now define two set functions

$$\underline{\mu}_{\kappa}(A) = \max\{\sum_{x\in A} \underline{\psi}^{k}(x), \kappa(\mathcal{X}) - \sum_{x\notin A} \overline{\psi}^{k}(x)\}, \tag{12}$$

$$\overline{\mu}_{\kappa}(A) = \min\{\sum_{x\in A} \overline{\psi}^{k}(x), \kappa(\mathcal{X}) - \sum_{x\notin A} \underline{\psi}^{k}(x)\}, \tag{13}$$

with the subscript $\kappa$ denoting that they come from an initial sub-modular set function. We will now proceed to show that $\underline{\mu}_{\kappa}$ and $\overline{\mu}_{\kappa}$ approximate $\kappa$ and that they are respectively sub-and super-modular. Indeed, even if $\kappa$ has a non-empty core $\mathcal{C}_{\kappa}$, we are not especially interested in positioning $\underline{\mu}_{\kappa}$ and $\overline{\mu}_{\kappa}$ with respect to this core, but with respect to the initial $\kappa$. Let us first show that $\underline{\mu}_{\kappa}$ and $\overline{\mu}_{\kappa}$ can act as lower/upper bounds.

**Proposition 5.** $\underline{\mu}_{\kappa} \leq \kappa^{c} \leq \overline{\mu}_{\kappa}$.

*Proof.* Let us first prove that $\underline{\mu}_{\kappa} \leq \kappa^{c}$. There are two possible values for $\underline{\mu}_{\kappa}(A)$:

- If $\underline{\mu}_{\kappa}(A) = \sum_{x\in A} \underline{\psi}^{k}(x)$ then $\underline{\mu}_{\kappa}(A) \leq \kappa^{c}(A)$ since $\kappa$ is a sub-modular set function.
- If $\underline{\mu}_{\kappa}(A) = \kappa(\mathcal{X}) - \sum_{x\notin A} \overline{\psi}^{k}(x)$ then $\sum_{x\notin A} \overline{\psi}^{k}(x) \geq \kappa(A^{c}) = \kappa(\mathcal{X}) - \kappa^{c}(A)$ which entails $\underline{\mu}_{\kappa}(A) \leq \kappa(\mathcal{X}) - \kappa(\mathcal{X}) + \kappa^{c}(A) = \kappa^{c}(A)$.

The proof of $\kappa^{c} \leq \overline{\mu}_{\kappa}$ is similar.

Note that Proposition 1 also applies here, from which follows that $\underline{\mu}_{\kappa}$ and $\overline{\mu}_{\kappa}$ can also be used to approximate $\kappa$, as we can reproduce Proposition 5 proof for $\kappa$.

**Corollary 1.** $\underline{\mu}_{\kappa} \leq \kappa^{c} \leq \kappa \leq \overline{\mu}_{\kappa}$

We can now prove that $\underline{\mu}_{\kappa}$ enjoys the same interesting property as the function $\kappa$ it approximates.

**Proposition 6.** *$\overline{\mu}_{\kappa}$ is a sub-modular set function i.e.,*

$$\forall A, B \subseteq \mathcal{X}, \overline{\mu}_{\kappa}(A\cup B) + \overline{\mu}_{\kappa}(A\cup B) \leq \overline{\mu}_{\kappa}(A) + \overline{\mu}_{\kappa}(B).$$

*Proof.* The proof is quite simple once we notice that $\mathcal{C}_{\kappa}$ is not empty and that

$$\underline{\psi}^{\kappa}(x) = \inf_{\psi\in\mathcal{C}_{\kappa}} \psi(x)$$
$$\overline{\psi}^{\kappa}(x) = \sup_{\psi\in\mathcal{C}_{\kappa}} \psi(x),$$

which can be obtained by applying Equations (9) and (7) to the indicator function $\mathbb{I}_{x}$ ($= 1$ on $x$ and zero everywhere else). Since the bounds $\underline{\psi}^{\kappa}$ and $\overline{\psi}^{\kappa}$ are atomic bounds over the set of additive measures $\mathcal{C}_{\kappa}$, one can apply Proposition 3.

**Corollary 2.** *$\overline{\mu}_\kappa$ is super-modular.*

Our results demonstrate that $\underline{\mu}_\kappa$ and $\overline{\mu}_\kappa$ are interesting approximations of a sub-modular function $\kappa$, possibly partially unknown. They are tight outer approximations, in the sense that as $\underline{\psi}^\kappa(x) = \kappa^c(\{x\})$ and $\overline{\psi}^\kappa(x) = \kappa(\{x\})$, any smaller intervals would not be outer approximations on the singletons. Finally, we can show an interesting link between the precision of the intervals $[\underline{\psi}^\kappa, \overline{\psi}^\kappa]$ and the functions $\overline{\mu}_\kappa, \underline{\mu}_\kappa$.

**Proposition 7.** *For all $A \subseteq \mathcal{X}$,*

$$|\overline{\mu}_\kappa(A) - \underline{\mu}_\kappa(A)| \leq card(A)||u - l||_{A,\infty}$$

*where $||u - l||_{A,\infty} = \max_{x \in A} |\overline{\psi}(x) - \underline{\psi}(x)|$ and $card(A)$is the cardinality of $A$.*

*Proof.* According to the definition of $\underline{\mu}_\kappa$ and $\overline{\mu}_\kappa$, we have for possible cases.
$\overline{\mu}_\kappa(A) - \underline{\mu}_\kappa(A)$ =

- $\overline{\mu}_\kappa(A) = \sum_{x \in A} \overline{\psi}^k(x) \leq \kappa(\mathcal{X}) - \sum_{x \notin A} \underline{\psi}^k(x)$
  $\underline{\mu}_\kappa(A) = \sum_{x \in A} \underline{\psi}^k(x) \geq \kappa(\mathcal{X}) - \sum_{x \notin A} \overline{\psi}^k(x)$.
  $\overline{\mu}_\kappa(A) - \underline{\mu}_\kappa(A) = \sum_{x \in A} (\overline{\psi}^k(x) - \underline{\psi}^k(x))$.
- $\overline{\mu}_\kappa(A) = \sum_{x \in A} \overline{\psi}^k(x) \leq \kappa(\mathcal{X}) - \sum_{x \notin A} \underline{\psi}^k(x)$
  $\underline{\mu}_\kappa(A) = \kappa(\mathcal{X}) - \sum_{x \notin A} \overline{\psi}^k(x) \geq \sum_{x \in A} \underline{\psi}^k(x)$.
  $\overline{\mu}_\kappa(A) - \underline{\mu}_\kappa(A) = \sum_{x \in A} \overline{\psi}^k(x) - \kappa(\mathcal{X}) + \sum_{x \notin A} \overline{\psi}^k(x) \leq \sum_{x \in A} (\overline{\psi}^k(x) - \underline{\psi}^k(x))$.
- $\overline{\mu}_\kappa(A) = \kappa(\mathcal{X}) - \sum_{x \notin A} \underline{\psi}^k(x) \leq \sum_{x \in A} \overline{\psi}^k(x)$
  $\underline{\mu}_\kappa(A) = \sum_{x \in A} \underline{\psi}^k(x) \geq \kappa(\mathcal{X}) - \sum_{x \notin A} \overline{\psi}^k(x)$.
  $\overline{\mu}_\kappa(A) - \underline{\mu}_\kappa(A) = \kappa(\mathcal{X}) - \sum_{x \notin A} \underline{\psi}^k(x) - \sum_{x \in A} \underline{\psi}^k(x) \leq \sum_{x \in A} (\overline{\psi}^k(x) - \underline{\psi}^k(x))$.
- $\overline{\mu}_\kappa(A) = \kappa(\mathcal{X}) - \sum_{x \notin A} \underline{\psi}^k(x) \leq \sum_{x \in A} \overline{\psi}^k(x)$
  $\underline{\mu}_\kappa(A) = \kappa(\mathcal{X}) - \sum_{x \notin A} \overline{\psi}^k(x) \geq \sum_{x \in A} \underline{\psi}^k(x)$.
  $\overline{\mu}_\kappa(A) - \underline{\mu}_\kappa(A) = \kappa(\mathcal{X}) - \sum_{x \notin A} \underline{\psi}^k(x) - \kappa(\mathcal{X}) + \sum_{x \notin A} \overline{\psi}^k(x) \leq \sum_{x \in A} (\overline{\psi}^k(x) - \underline{\psi}^k(x))$.

So we have $|\overline{\mu}_\kappa(A) - \underline{\mu}_\kappa(A)| \leq card(A)||u - l||_{A,\infty}$.

Note that if $\kappa$ is linear the previous result proves that $\overline{\mu}_\kappa = \underline{\mu}_\kappa$ which is trivially true. More over, the further $\kappa$ is from additivity ,the greater the difference $|\overline{\mu}_\kappa(A) - \underline{\mu}_\kappa(A)|$ can be. This result also seems to connect our approximation with the total variation distance, routinely used in statistics [8].

Finally, let us remark that in Example 2, we would have derived the same bounds if we assumed that the group computed the utility of a subset $A$ according to a sub-modular aggregation function. In particular, this means that if we concentrate on atomic bounds only, it does not really matter whether $\underline{\mu}$ and $\overline{\mu}$ approximate a set $\mathcal{M}$ of additive set functions or a sub-modular set function (which itself can be considered as a non-empty set of dominated additive set functions), as both approximations will coincide if the bounds $\underline{\psi}, \overline{\psi}$ are the same.

# 4 An application to signal convolution

We illustrate our approximation technique on an application in signal processing, as signal processing is strongly based on the notion of linear system, i.e. additive set functions.

## 4.1 Signal processing and additive set functions

We speak of continuous systems when the signals are continuous and of discrete systems when the signals are discrete. However, a signal is rarely discrete by nature. Discrete signals are generally continuous signals that have been discretized. For example, natural digital images are discretizations, by the retina of a camera, of continuous luminance images in real space. When they are naturally discrete, signals are often produced to mimic a discretized signal (synthetic images and sounds) with the aim of being used in the real world (special effects, electronic music, etc.).

Except in very rare cases, signal processing only really makes sense in continuous space. When the processing of the sound signal is used to recognize a sound for example (voice dictation) it is indeed recognition of the continuous signal that is involved via the recognition of the discrete signal from which it originates. So, doing digital signal processing generally means digitally performing a modification or analysis of the continuous signal from which it is being sampled. We limit ourselves here to linear signal processing.

The classical view consists in supposing that the processing of the signal, for which it is a question of finding a digital equivalent, is obtained thanks to a known continuous linear system (e.g. noise reduction filter (sound), texture modification (images), ...).

When a system (continuous or discrete) is linear, then the relation between input and output is completely defined by a continuous ($\varphi$) or discrete ($\psi$) signal called *impulse response of the system* and a linear operation called *convolution*. If the system is continuous, $s$ the output of the system is written as a function of $e$ the input:

$$s(t) = (e * \varphi)(t) = \int_{-\infty}^{\infty} e(u).\varphi(t-u)du, \tag{14}$$

and in the discrete case :

$$s(y) = (e * \psi)(y) = \sum_{x=-\infty}^{\infty} e(x).\psi(y-x). \tag{15}$$

Most signal processing being inherently continuous, one needs to go from the continuous ideal impulse response $\varphi(t), t \in \mathbb{R}$ to a corresponding discrete impulse response $\psi(x)\ x \in \mathbb{Z}$. One standard way to do that is to consider $\psi(x) = \beta.\varphi(k.\Delta)$, $\Delta$ being the sampling period, and $\beta$ a normalization coefficient allowing to respect the dynamics of the system, i.e. $\int_{\mathbb{R}} \varphi(t)dt = \sum_{x\in\mathbb{Z}} \psi(x)$. This approach is based on a modeling of the sampling by a Dirac impulse, which is very far from reality.

A more reasonable way to see the problem is to assume that the sampling was performed by a real system allowing to go from continuous space to discrete space whose impulse response $\eta$ is of finite support and bound to the sampling period $\Delta$. This

is written $\psi(x) = (\varphi * \eta)(x.\Delta)$ (see Figure (1)). $\eta$ is called the summative *sampling kernel*, summative meaning that it integrates to one ($\int_{\mathbb{R}} \eta(t)dt = 1$) and its support is included in $[-\Delta, \Delta]$ (i.e. $\forall t \notin [-\Delta, \Delta]$, $\eta(t) = 0$).

Note that the input signal $e$ being generally finite, i.e. it is made of $n$ samples, the expression (15) is generally written:

$$s(y) = (e * \psi)(y) = \sum_{x=1}^{n} e(x).\psi(y - x) = \mathbb{E}_{\kappa_{\psi_y}}(e), \tag{16}$$

with $\forall y \in \mathbb{Z}$, $\psi_y(y) = \psi(y - x)$. When the operation of the equation (16) does not represent a system but a mathematical operation aimed at modifying a signal, $\psi$ is more generally named *convolution kernel*.

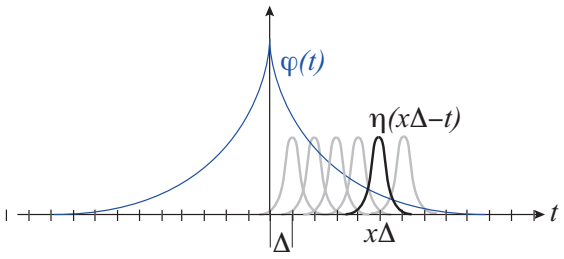


Fig. 1: Sampling of the impulse response $\varphi$ with the kernel $\eta$.

### 4.2 Sampling and fuzzy transformation

A major problem posed by this approach is that the values of $\psi$ strongly depend on the arbitrary choice of the chosen sampling kernel. For the discrete-continuous approach to be valid, the continuous convolution kernel must be digitized with the same sampling kernel as the discrete signal to be filtered. In most applications, this kernel is at best unknown or even, in the case of image processing, non-existent because the sampling is not invariant by translation: each pixel has its own impulse response.

To overcome this problem, Loquin et al. [13] propose to replace the summative sampling kernel by a family of continuous sampling kernels (which are formally equivalent to probability distributions) represented by a possibility distribution $\pi$ [9], and called maxitive kernels. Within the framework of this theory, the triangular maxitive kernel plays a particular role since it allows to represent all the symmetric monomodal sampling kernels of lesser granularity [2].

This technique has been used in [17] under the name of imprecise fuzzy transformation to obtain an interval-valued impulse response $[\underline{\psi}(x), \overline{\psi}(x)]$ ($x \in \mathbb{Z}$) of all the values that could have been obtained by considering a family of sampling kernels represented by $\pi$.

We propose to extend the convolution operation defined in Equation (16) to the interval-valued impulse response $[\underline{\psi}, \overline{\psi}]$ by:

$$[\underline{s}(y), \overline{s}(y)] = (e * [\underline{\psi}, \overline{\psi}])(y) = [\underline{\mathbb{E}}_{\underline{\mu}_{\mathcal{M}_y}}(e), \overline{\mathbb{E}}_{\overline{\mu}_{\mathcal{M}_y}}(e)], \tag{17}$$

with $\underline{\mu}_{\mathcal{M}_y}$ being defined by Equations (10), considering $\forall x \in \mathbb{Z}$, $\underline{\psi}_y(x) = \underline{\psi}(y - x)$ and likewise for $\overline{\mu}_{\mathcal{M}_y}$.

### 4.3 Derivation of a discrete signal

The derivation of a discrete signal does not make sense. Indeed, only continuous signals can be differentiable. However, many signal analysis techniques are based on their derivation (extraction of contours or particular points in images, extraction of patterns in physiological signals, ...). In this context, it is not the discrete signal that we want to derive. What we want to do is to find an operation equivalent to a derivation of the continuous signal (before digitization) then sampling of the derivative obtained. It has been shown in [21] that this succession of operations is equivalent to convoluting the discrete signal with the sampling of the derivative of the kernel used for the reconstruction. One of the kernels often used to perform this operation is the Shen-Castan kernel [16]

$$\forall t \in \mathbb{R}, \quad \phi(t) = -\frac{\ln(\beta)}{2}.\beta^{|t|}, \tag{18}$$

where $\beta \in ]0, 1[$ is a form factor of the kernel. The derivation of the kernel $\phi$ leads to the continuous kernel

$$\forall t \in \mathbb{R}, \quad \varphi(t) = -\operatorname{sign}(t).\frac{\ln(\beta)^2}{2}.\beta^{|t|}$$

Let us assume, without any loss of generality, that the sampling step $\Delta = 1$. Then, sampling $\varphi$ as proposed in Section 4.2 by using a possibility distribution $\pi$ leads to the interval-valued impulse response of the form $[\underline{\psi}, \overline{\psi}]$:

$$\begin{aligned}
\text{if } x > 0,\ \overline{\psi}(x) &= -\frac{\ln(\beta).(\beta^{|x+1|} - \beta^{|x|})}{2}, \\
\underline{\psi}(x) &= -\frac{\ln(\beta).(\beta^{|x|} - \beta^{|x-1|})}{2}, \\
\text{and if } x < 0,\ \overline{\psi}(x) &= \frac{\ln(\beta).(\beta^{|x|} - \beta^{|x-1|})}{2}, \\
\underline{\psi}(x) &= \frac{\ln(\beta).(\beta^{|x|} - \beta^{|x|})}{2}.
\end{aligned}$$

It is interesting to note that, in this application, $\underline{\mu}_{\mathcal{M}_y}(\mathcal{X}) = \overline{\mu}_{\mathcal{M}_y}(\mathcal{X}) \approx 0$. Figure (2) shows the corresponding bounds, together with some discrete $\psi$ obtained by the use of sampling kernels that are within the family represented by $\pi$, that is symmetric kernels with support $[-\delta, \delta]$ ($\delta \in ]0, 1]$). Such kernels includes, e.g., linear and quadratic kernels, as well as the Dirac kernel (whose result is in black in Figure (2)).

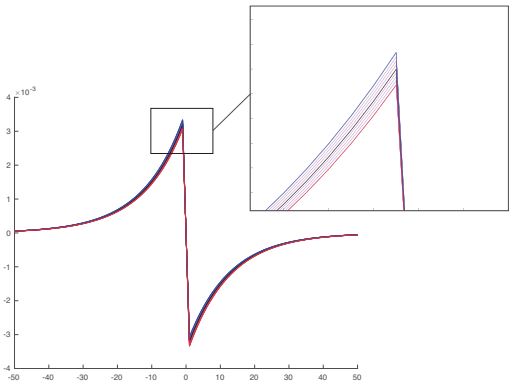


Fig. 2: $\overline{\psi}$ (in blue), $\underline{\psi}$ (in red), 8 impulse responses for different kernels within the family represented by $\pi$ (in magenta), and the response of the Dirac kernel (in black).

Figure (3) and Figure (4) respectively show a time-attenuated signal and its derivative bounds obtained by our approach, together with some sampled filters. We can see on Figure (4) that our approach makes a good job, providing wide bounds when variations are important (bottom box in Figure (4)), while providing very tight bounds when variation decreases along time (upper box).

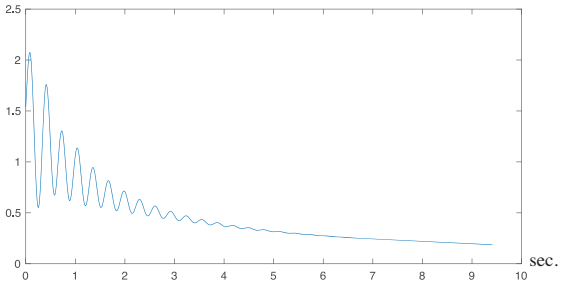


Fig. 3: Synthesis signal composed of a sum of attenuated sinusoids.

## 5 Conclusion

In this paper, we have discussed how having only atomic bounds can be used to efficiently approximate very generic sets of very generic additive or sub-modular measures (i.e., non-monotonic with arbitrary bounds).

We have shown how they can successfully be applied to perform robust signal processing. Our next step will be to investigate how our results regarding sub-modular and

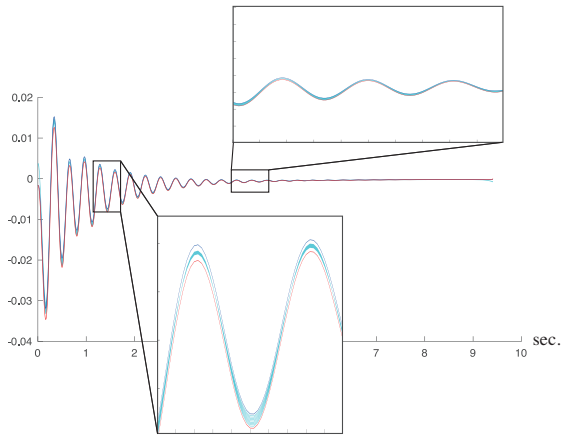


Fig. 4: Original signal derived by the imprecise Shen filter (lower bound in red, upper bound in blue) superimposed with derived signals obtained by various sampled Shen filters.

super-modular set functions can be instrumental to represent non-monotonic rankings of sets or as approximation tools in optimisation problems using sub-modular functions.

## References

1. M. Cohen A. Chateauneuf. Cardinal extensions of the eu model based on the choquet integral. In D. Bouyssou et al., editor, *Decision-making Process – Concepts and Methods*, chapter 9, pages 401–434. Wiley, 2009.
2. Cédric Baudrit and Didier Dubois. Practical representations of incomplete probabilistic knowledge. *Computational statistics & data analysis*, 51(1):86–108, 2006.
3. Nawal Benabbou, Cassandre Leroy, Thibaut Lust, and Patrice Perny. Interactive optimization of submodular functions under matroid constraints. In *International Conference on Algorithmic Decision Theory*, pages 307–322. Springer, 2021.
4. H. Prade D. Dubois. *Possibility Theory. An Approach to Computerized Processing of Uncertainty*. New York and London, 1988.
5. H. Prade D. Dubois. *Possibility theory and its applications: Where do we stand?* Springer, 2015.
6. Luis M De Campos, Juan F Huete, and Serafin Moral. Probability intervals: a tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2(02):167–196, 1994.
7. Anja De Waegenaere, Peter Wakker, et al. *Choquet integrals with respect to non-monotonic set functions*. Center for Economic Research, Tilburg University, 1997.
8. Luc Devroye and Laszlo Gyorfi. No empirical probability measure can converge in the total variation sense for all distributions. *The Annals of Statistics*, pages 1496–1499, 1990.
9. Didier Dubois and Henri Prade. When upper probabilities are possibility measures. *Fuzzy sets and systems*, 49(1):65–74, 1992.
10. Peter C Fishburn and Irving H LaValle. Binary interactions and subset choice. *European journal of operational research*, 92(1):182–192, 1996.

11. Victor Gabillon, Branislav Kveton, Zheng Wen, Brian Eriksson, and Shanmugavelayutham Muthukrishnan. Adaptive submodular maximization in bandit setting. In *NIPS*, pages 2697–2705. Citeseer, 2013.
12. M. Grabisch. *Set functions, Games and Capacities in Decision-Making*. 2016.
13. Kevin Loquin and Olivier Strauss. On the granularity of summative kernels. *Fuzzy sets and systems*, 159(15):1952–1972, 2008.
14. Stefano Moretti and Meltem Öztürk. Some axiomatic and algorithmic perspectives on the social ranking problem. In *International Conference on Algorithmic DecisionTheory*, pages 166–181. Springer, 2017.
15. Alexander Schrijver et al. *Combinatorial optimization: polyhedra and efficiency*, volume 24. Springer, 2003.
16. Jun Shen and Serge Castan. Towards the unification of band-limited derivative operators for edge detection. *Signal Processing*, 31(2):103–119, 1993.
17. Olivier Strauss. Non-additive interval-valued f-transform. *Fuzzy sets and systems*, 270:1–24, 2015.
18. Ehsan Tohidi, Rouhollah Amiri, Mario Coutino, David Gesbert, Geert Leus, and Amin Karbasi. Submodularity in action: From machine learning to signal processing applications. *IEEE Signal Processing Magazine*, 37(5):120–133, 2020.
19. Sebastian Tschiatschek, Adish Singla, and Andreas Krause. Selecting sequences of items via submodular maximization. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
20. Naoya Uematsu, Shunji Umetani, and Yoshinobu Kawahara. An efficient branch-and-cut algorithm for submodular function maximization. *Journal of the Operations Research Society of Japan*, 63(2):41–59, 2020.
21. Michael Unser, Akram Aldroubi, and Murray Eden. B-spline signal processing. i. theory. *IEEE transactions on signal processing*, 41(2):821–833, 1993.
22. L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1978.