Smooth Non-Additive Integrals and Measures and Their Potential Applications

Olga Kosheleva  
*The University of Texas at El Paso*, olgak@utep.edu

Vladik Kreinovich  
*The University of Texas at El Paso*, vladik@utep.edu

Follow this and additional works at: [https://scholarworks.utep.edu/cs_techrep](https://scholarworks.utep.edu/cs_techrep)  
Part of the Computer Sciences Commons, and the Mathematics Commons

Comments:  
Technical Report: UTEP-CS-23-68

Recommended Citation  
[https://scholarworks.utep.edu/cs_techrep/1853](https://scholarworks.utep.edu/cs_techrep/1853)

This Article is brought to you for free and open access by the Computer Science at ScholarWorks@UTEP. It has been accepted for inclusion in Departmental Technical Reports (CS) by an authorized administrator of ScholarWorks@UTEP. For more information, please contact Iweber@utep.edu.
Smooth Non-Additive Integrals and Measures and Their Potential Applications

Olga Kosheleva and Vladik Kreinovich

Abstract In this paper, we explain why non-additive integrals and measures are needed, how non-additive integrals and measures are related, how to use them in decision making, and how they can help in fundamental physics. These four topics are covered, correspondingly, in Sections 2–5 of this paper.

1 Additive Integrals and Measures: What Are They And Why They Are Needed: Reminder

How to best influence an area: practical need. In many practical situations, we need to influence a 2-D or 3-D area. For example:

• in medicine, we want to affect all the locations of a tumor;
• in humanitarian aid situations, we need to affect all the locations in the suffering area;
• in communications – e.g., in a health campaign or in a political campaign – we want to make sure that the desired message reaches all the people in a given area;
• in a military attack, we want to disable as many adversary troops as possible, etc.

In many such situations, our abilities to influence are limited. So, it is important to make sure that we reach the largest possible effect by using the given resources.

How can we describe this practical problem in precise terms: general idea. Let $R$ denote the overall amount of resources. The problem is to how to distribute
these resources over a given area $A$. This distribution can be described by a density $f(x) \geq 0$, so that the overall amount of resources is $R$:

$$\int_A f(x) \, dx = R. \quad (1)$$

The overall effect $E$ of such distribution depends on all the values $f(x)$. In many practical situations, the dependence is smooth, in the sense that small changes in $f(x)$ lead to small changes in effect. In such situations, a reasonable idea is to expand the dependence of $E$ on the values $f(x)$ in Taylor series, i.e., consider the expression

$$E = a_0 + \int_A a_1(x) \cdot f(x) \, dx + \int_A \int_A a_2(x, x') \cdot f(x) \cdot f(x') \, dx \, dx' + \int_A \int_A \int_A a_3(x, x', x'') \cdot f(x) \cdot f(x') \cdot f(x'') \, dx \, dx' \, dx'' + \ldots \quad (2)$$

for some symmetric functions $a_i$ — i.e., we have $a_2(x, x') = a_2(x', x)$, $a_3(x, x', x'') = a_2(x', x, x'')$, etc.

**Comments.**

- Here, we have infinitely many locations $x$ and thus, infinitely many quantities $f(x)$. So, instead of the usual sum of terms corresponding to different variables, we have an integral — the limit of such sums.
- If we do not do anything, i.e., if we take $f(x) = 0$ for all $x$, we will have no effect at all, i.e., we will have $E = 0$. Substituting $f(x) = 0$ into the general formula (1), we conclude that $a_0 = 0$ and thus, the formula (2) takes the following form:

$$E = \int_A a_1(x) \cdot f(x) \, dx + \int_A \int_A a_2(x, x') \cdot f(x) \cdot f(x') \, dx \, dx' + \int_A \int_A \int_A a_3(x, x', x'') \cdot f(x) \cdot f(x') \cdot f(x'') \, dx \, dx' \, dx'' + \ldots \quad (3)$$

**Case of small influence.** We consider situations in which our abilities are limited, i.e., when the overall amount of resources $R$ is small. In this case, it seems reasonable to assume that all the values $f(x)$ are also reasonably small. Since the values $f(x)$ are small, terms which are quadratic (or higher order) in terms of $f(x)$ are much smaller than linear terms and can, therefore, be safely ignored. For example, if $f(x)$ is about 1%, its square is 0.01% – much much smaller.

If we ignore all nonlinear terms in the general expression (3), we get the following formula for the overall effect:

$$E = \int_A a_1(x) \cdot f(x) \, dx. \quad (4)$$
**Description in terms of measures and integrals.** In particular, for each set \( S \), for the function \( f(x) = \chi_S(x) \) for which \( \chi_S(x) = 1 \) for \( x \in S \) and \( \chi_S(x) = 0 \) for \( x \notin S \), the formula (4) leads to
\[
E = \int_S a_1(x) \, dx.
\] (5)

This value is often denoted by \( \mu(S) \).

The function \( \mu(S) \) that assigns this value to each measurable set \( S \) is called **additive measure**, or simply **measure**, for short. The word “additive” reflects the following property of this function: if the two sets \( S \) and \( S' \) do not have any common elements, then
\[
\mu(S \cup S') = \mu(S) + \mu(S').
\] (6)

The expression (5) is often written in the following form:
\[
E = \int f(x) \, d\mu.
\] (7)

This is called an **integral** of the function \( f(x) \) over the measure \( \mu \).

**We can reconstruct the measure if we know the integral, and vice versa.** To present the corresponding result, let us start with definitions.

**Definition 1.** Let \( a_1(x) \) be a continuous function.

- By an additive integral, we mean a function of type (4) that transforms a function \( f(x) \) into the value \( E(f) \).
- By an additive measure corresponding to an additive integral \( E(f) \), we mean a function that assigns, to each measurable set \( S \subseteq A \), the value \( \mu(S) \) defined \( E(\chi_S) \).

**Proposition 1.**

- Once we know the additive integral, we can uniquely reconstruct the corresponding additive measure.
- Vice versa, once we know the additive measure, we can uniquely reconstruct the corresponding additive integral.

**Proof.**

1°. We have already shown that once we know the integral, we can uniquely reconstruct the measure: namely, \( \mu(S) = \int_\chi_S(x) \, d\mu \).

2°. Let us now prove that, vice versa, if we know the measure, then we can reconstruct the corresponding function \( a_1(x) \) and thus, uniquely reconstruct the corresponding additive integral.

Indeed, for each point \( x_0 \), we can form the set of all the points \( x \) whose distance from \( x_0 \) does not exceed \( \varepsilon \): \( B_\varepsilon(x_0) = \{x : d(x, x_0) \leq \varepsilon\} \). This set is called a **ball** of radius \( \varepsilon \) with a center at \( x_0 \). For continuous functions \( a_1(x) \), we have
\[
\mu(B_\varepsilon(x_0)) = \int_{B_\varepsilon(x_0)} a_1(x) \, dx = a_1(x_0) \cdot V(\varepsilon) + o(V(\varepsilon)),
\] (7)
where \( V(\varepsilon) \) denotes the volume of a ball of radius \( \varepsilon \). Thus, we can find \( a_1(x_0) \) as

\[
a_1(x_0) = \lim_{\varepsilon \to 0} \frac{\mu(B_\varepsilon(x_0))}{V(\varepsilon)}.
\]

(8)

The proposition is proven.

So what is the optimal way to influence: seemingly natural answer. Now that we have a description of the effect of different distributions, we can start answering the original question:

• given the limited resources, i.e., given the constraint (1),
• how can reach the largest possible effect, i.e., the largest possible value of the integral (4)?

The answer to this question is as follows.

**Proposition 2.** To maximize the expression (4) under the constraint (1), we need to concentrate \( f(x) \) on on the point(s) \( x_0 \) at which the value \( a_1(x) \) is the largest, i.e., at which

\[
a_1(x_0) = M \overset{\text{def}}{=} \max_x a_1(x),
\]

i.e., we should take \( f(x) = 0 \) for all other points.

**Proof.** Indeed, if we make this concentration, then in all the points in which \( f(x) \neq 0 \), we have \( f(x) = M \). Thus, for this concentration, we have

\[
E = \int a_1(x) \cdot f(x) \, dx = \int a_1(x) \cdot M \, dx = M \cdot \int a_1(x) \, dx = M \cdot R.
\]

On the other hand, for every other distribution \( f(x) \), we have \( a_1(x) \leq M \) for all \( x \), and we have \( a_1(x_1) < M \) for some point \( x_1 \) for which \( f(x_1) > 0 \). Due to continuity, we have \( a_1(x) < M \) for all \( x \) for some vicinity \( V \) of \( x_1 \). Thus, we have

\[
E = \int a_1(x) \cdot f(x) \, dx < \int a_1(x) \cdot M \, dx = M \cdot \int a_1(x) \, dx = M \cdot R.
\]

Thus, for these distributions, the functional \( E(f) \) does not attain its largest value. The proposition is proven.

2 Why We Need Non-Additive Integrals and Measures

**What is a problem with the above optimal solution.** The problem with this seemingly natural solution is that we concentrate all the efforts in a single location – i.e., in effect, on a very small area of small volume \( V \). In this case, the value \( f(x_0) = R/V \) becomes large.

And for large values, we can no longer ignore quadratic and higher order terms in the general expression (3).
Conclusion: we need to go beyond additive measures and integrals. Because of the above problem, we need to consider general expression (3) instead of a simplified additive expression (4).

Comment. The need to use non-additive measures was emphasized in many papers by Michio Sugeno – to whose memory this book is dedicated – and his students and colleagues; see, e.g., [5].

3 How Non-Additive Integrals and Measures Are Related

In the first section, we showed that:

- once we have a usual (additive) integral, we can uniquely reconstruct the corresponding additive measure, and
- vice versa, once we have an additive measure, we can uniquely reconstruct the corresponding integral.

In this section, we show that a similar relation holds in the general – not necessary additive – case. Let us start with the definitions.

**Definition 2.** Let $a_1(x)$, $a_2(x, x')$, $a_3(x, x', x'')$, ... be continuous functions.

- By a non-additive integral, we mean a functional of type (3) that transforms a function $f(x)$ into a value $E(f)$.
- By a non-additive measure corresponding to a non-additive integral $E(f)$, we mean a function that assigns, to each measurable set $S \subseteq A$, the value $\mu(S) \overset{\text{def}}{=} E(\chi_S)$.

**Proposition 3.**

- Once we know the non-additive integral, we can uniquely reconstruct the corresponding non-additive measure.
- Vice versa, once we know the non-additive measure, we can uniquely reconstruct the corresponding non-additive integral.

**Proof.**

1°. Once we know the integral, we can uniquely reconstruct the measure: namely, $\mu(S) = \int \chi_S(x) \, dx$.

2°. Let us now prove that, vice versa, if we know the measure, then we can reconstruct the corresponding functions $a_1(x)$, $a_2(x, x')$, $a_3(x, x', x'')$, etc., and thus, uniquely reconstruct the corresponding additive integral.

2.1°. Indeed, in this case, we still have

$$\mu(B_\varepsilon(x_0)) = \int_{B_\varepsilon(x)} a_1(x) \, dx + o(V(\varepsilon)) = a_1(x_0) \cdot V(\varepsilon) + o(V(\varepsilon)). \quad (9)$$
Thus, we can find $a_1(x_0)$ as

$$a_1(x_0) = \lim_{\varepsilon \to 0} \frac{\mu(B_\varepsilon(x_0))}{V(\varepsilon)}. \quad (10)$$

2.2°. Once we have found the function $a_1(x)$, we can now form an auxiliary measure

$$\mu_2(S) \overset{\text{def}}{=} \mu(S) - \int_S a_1(x) \, dx = \int_S \int_S a_2(x, x') \, dx \, dx' + \ldots \quad (11)$$

For this functional, we have

$$\mu_2(B_\varepsilon(x_0)) = a_2(x_0, x_0) \cdot (V(\varepsilon))^2 + o((V(\varepsilon))^2). \quad (12)$$

Thus, we can determine the values $a_2(x, x)$ as the limit:

$$a_2(x_0, x_0) = \lim_{\varepsilon \to 0} \frac{\mu_2(B_\varepsilon(x_0))}{(V(\varepsilon))^2}. \quad (13)$$

Now, for every two points $x_0$ and $x'_0$, we have

$$\mu_2(B_\varepsilon(x_0) \cup B_\varepsilon(x'_0)) = a_2(x_0, x_0) \cdot (V(\varepsilon))^2 + a_2(x'_0, x'_0) \cdot (V(\varepsilon))^2 +$$

$$2a_2(x_0, x'_0) \cdot (V(\varepsilon))^2 + o((V(\varepsilon))^2). \quad (14)$$

Thus, in the limit, we can determine the sum

$$a_2(x_0, x_0) + a_2(x'_0, x'_0) + 2a_2(x_0, x'_0) = \lim_{\varepsilon \to 0} \frac{\mu_2(B_\varepsilon(x_0) \cup B_\varepsilon(x'_0))}{(V(\varepsilon))^2}. \quad (15)$$

Since we already know the values $a_2(x_0, x_0)$ and $a_2(x'_0, x'_0)$, we can therefore determine the value $a_2(x_0, x'_0)$ for all possible pairs $(x_0, x'_0)$.

2.3°. Once we have found the function $a_2(x, x')$, we can now form an auxiliary measure

$$\mu_3(S) \overset{\text{def}}{=} \mu_2(S) - \int_S \int_S a_2(x, x') \, dx \, dx' =$$

$$\int_S \int_S \int_S a_3(x, x', x'') \, dx \, dx' \, dx'' + \ldots \quad (16)$$

Similarly to Part 2.2 of this proof, we can now:

- first, determine the values $a_3(x, x, x)$,
- then the values $a_3(x, x, x')$ for all pairs $(x, x')$, and
- finally, the values $a_3(x, x', x'')$ for all triples $(x, x', x'')$.

Once we have determined the function $a_3$, we can subtract all cubic terms from $\mu_3(S)$ and thus, get an expression that starts with $a_4$. From this expression, we can similarly determine $a_4$, etc.
The proposition is proven.

4 How to Use Non-Additive Integrals and Measures in Decision Making

Let us consider the simplest non-linear case. Our main motivation for using non-additive integrals and measures was that for decision making, the linear approximation – corresponding to additive integrals and measures – may be too crude, and a more accurate approximation is needed. The next approximation after linear is quadratic, when

\[ E = \int_A a_1(x) \cdot f(x) \, dx + \int_A \int_A a_2(x, x') \cdot f(x) \cdot f(x') \, dx \, dx'. \]  

(17)

Resulting optimization problem. We want to maximize the quadratic expression (17) under the constraint (1).

Additional assumption. Let us make a reasonable assumption that \( f(x) > 0 \) for all \( x \).

How to solve the corresponding optimization problem: analysis. For this problem, the usual Lagrange multiplier method for solving constraint optimization problems means that we reduce this problem to an unconstrained optimization problem of maximizing the expression

\[ E + \lambda \cdot \left( \int_A f(x) \, dx - R \right) = \int_A a_1(x) \cdot f(x) \, dx + \int_A \int_A a_2(x, x') \cdot f(x) \cdot f(x') \, dx \, dx' + \lambda \cdot \left( \int_A f(x) \, dx - R \right), \]  

(18)

for some constant \( \lambda \). This constant needs to be computed based on the condition (1).

Differentiating the expression (18) by each unknown \( f(x) \) and equating the derivative to 0, we conclude that

\[ a_1(x) + 2 \int a_2(x, x') \cdot f(x') \, dx' + \lambda = 0, \]  

(19)

i.e., equivalently,

\[ 2 \int a_2(x, x') \cdot f(x') \, dx' = -a_1(x) - \lambda. \]  

(20)

This equation is linear in terms of \( f(x) \). Thus, it is sufficient to solve the following two linear problems:
• find the function $f_1(x)$ for which

$$2 \int a_2(x, x') \cdot f_1(x') \, dx' = -a_1(x), \quad (20)$$

• and find the function $f_2(x)$ for which

$$2 \int a_2(x, x') \cdot f_2(x') \, dx' = -1. \quad (21)$$

In terms of these functions, we have $f(x) = f_1(x) + \lambda \cdot f_2(x)$. The condition (1) then takes the form

$$\int_A f_1(x) \, dx + \lambda \cdot \int_A f_2(x) \, dx = R, \quad (22)$$

due to which we can determine $\lambda$ as follows:

$$\lambda = \frac{R - \int_A f_1(x) \, dx}{\int_A f_2(x) \, dx}. \quad (23)$$

As a result, we arrive at the following algorithm.

**Resulting algorithm.** To solve the desired constraint optimization problem, we:

• first solve two linear systems (20) and (21), resulting in functions $f_1(x)$ and $f_2(x),$

• then, we compute $\lambda$ by using the formula (23), and

• finally, we compute $f(x) = f_1(x) + \lambda \cdot f_2(x)$.

## 5 How Non-Additive Integrals and Measures Can Help in Fundamental Physics

**What is the energy of the gravitational field — a known paradox.** In physics, there is a known paradox related to the energy of the gravitational field:

• on the one hand, gravitational forces are definitely perform useful work — this is, for example, how hydroelectric stations work, they transform the energy of the gravitational field into electricity;

• on the other hand, if we use the usual field-theoretic ideas to compute the energy of the gravitational field, we get zero.

To explain this paradox, let us recall how in physics, energy of different fields is determined.

**Where does this paradox come from: a brief reminder.** In physics, dynamic equations describing how the world changes usually follow from the so-called minimal action principle (see, e.g., [1, 2, 3, 4]), according to which the fields should minimize the “action” $S = \int L \, dx$, where the integrated expression $L$ — called Lagrangian — depends on the fields and their derivatives.
Similarly to the usual calculus, to find the corresponding minimum, we differentiate the minimized expression $S$ over all the components of the fields, and equate the resulting derivative to 0. When we differentiate a functional depending on the unknown function $f$ and on its derivatives, this differentiation is called \textit{variational differentiation}. Variational derivative has a special notation $\frac{\delta L}{\delta f}$, so the corresponding physical equations have the form

$$\frac{\delta L}{\delta f} = 0.$$  \hfill (24)

In general relativity, geometry of space-time is considered one of the fields. To be more precise, the corresponding field $g^{i,j}$, $(0 \leq i, j \leq 3)$ describes the proper time $ds$ along the small trajectory in which all 4 coordinates (including time $x_0$) change by $dx$: namely, $ds^2 = \sum_{i,j} g^{i,j} \cdot dx_i \cdot dx_j$. Interestingly, it turns out that for almost all fields, if we consider its Lagrangian $L$ in the general curved space-time, then the expression $T_{i,j} \overset{\text{def}}{=} \frac{\delta L}{\delta g^{i,j}}$, which is known as the \textit{energy-momentum tensor}, describes this field’s energy. Specifically, the overall energy of the field – i.e., its ability to perform useful work – is equal to $\int T_{0,0} \, dx_1 \, dx_2 \, dx_3$, so that the value $T_{0,0}$ describes the energy density.

This works for almost all fields, but the important exception is the gravitational field itself. Indeed, for this field, the corresponding equations (24) have exactly the form

$$\frac{\delta L}{\delta g^{i,j}} = 0.$$  \hfill (25)

In other words, the energy of gravitational field – as computed by the usual formulas of field theory – is 0.

\textbf{How non-additive integrals and measures can help.} There are two facts:

- the density of the gravitational field’s energy is 0, and
- the overall energy of the gravitational field is not 0.

These two facts lead to a paradox if we assume – as it is often done – that the overall energy $E(S)$ of the gravitation field in a region $S$ is equal to the integral of energy density $\rho(x)$ over this region, i.e., if $E(S) = \int_S \rho(x) \, dx$. In other words, this becomes a paradox if we assume that energy is described by an \textit{additive} measure.

However, the very fact that the density is 0 while the overall energy is not 0 shows that energy is not described by an additive measure – and thus, it can only be adequately described by a \textit{non-additive} measure.

\textbf{What is the physical meaning of this?} But what is the physical meaning of this phenomenon? This physical meaning – that energy is not localized, that the overall
energy of a system may not be equal to the sum of energies of its subregions – can be illustrated on a very simple example.

Suppose that in an empty area of the Universe, we have a stable particle that it moving with a constant speed in the same direction. In some reference frame, this particle is stationary. There is no way to make this particle, by itself, do any useful work. (We could if this was an unstable particle, we could use the products of its decay, but we assumed that the particle is stable.) From this viewpoint, its energy is 0 – and since this is the only particle in this area, the overall energy in this area is also 0.

Suppose now that in a neighboring area there is a similar stable particle that is also moving with a constant speed in the same direction – but either its speed or its direction are different from the first particle. In this case, the energy of the area containing this second particle is also 0. However, when we consider the union of these two areas, we do have an ability to perform useful work: namely, if these two particle collide, we can get energy out of it and thus, perform some work.

Acknowledgments

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), HRD-1834620 and HRD-2034030 (CAHSI Includes), EAR-2225395 (Center for Collective Impact in Earthquake Science C-CIES), and by the AT&T Fellowship in Information Technology.

It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and by a grant from the Hungarian National Research, Development and Innovation Office (NRDI).

References