Introduction to Fuzzy and Possibilistic Optimization

Weldon A. Lodwick and Elizabeth Untiedt

1 Introduction

Deterministic optimization is a normative process which extracts the best from a set of options, usually under constraints. It is arguably true that optimization is one of the most used areas of mathematical applications. It is the thesis of this book that applied mathematical programming problems should be solved predominantly by using a fuzzy and possibilistic approaches. Rommelfanger ([42], p. 295), states that the only operations research methods that is widely applied is linear programming. He goes on to state that even though this is true, of the 167 production (linear) programming systems investigated and surveyed by Fandel [18], only 13 of these were "purely" (my interpretation) linear programming systems. Thus, Rommelfanger concludes that even with this most highly used and applied operations research method, there is a discrepancy between classical linear programming and what is applied. Deterministic and stochastic optimization models require well-defined input parameters (coefficients, right-hand side values), relationships (inequalities, equalities), and preferences (real valued functions to maximize, minimize) either as real numbers or real valued distribution functions. Any large scale model requires significant data gathering efforts. If the model has projections of future values, it is clear that real numbers and real valued distributions are inadequate representations of parameters, even assuming that the model correctly captures the underlying system. It is also known from mathematical programming theory that only a few of the variables and constraints are necessary to describe an optimal solution (basic variables and active constraints), assuming a correct deterministic normative criterion (objective function). The ratio of variables that are basic and constraints that are active compared to the total becomes smaller, in general, as the model increases in size since in general large-scale models tend to become more sparse. Thus, only a few parameters need to be obtained precisely. Of course the problem is that it is not known *a priori* which variables will be basic and which constraints will be active.

Moreover, a model of an actual problem is always an abbreviated view of the underlying actual system. If a problem is able to be manipulated *in situ* to obtain a solution without a symbolic representation, then there would be no need for modeling the problem mathematically. Inherently, a mathematical model is a symbolic representation of the problem. Bertrand Russell ([44], p. 85, 86) states,

> "The law of excluded middle is true when precise symbols are employed, but it is not true when symbols are vague, as, in fact, all symbols are."

At the heart of (deterministic) pure mathematics are theorems whose proofs are mathematical statements that are either true or false, but not both (law of the excluded middle). This concept is also stated as a property of classical sets - either an element belongs to a set or does not, but not both. Fuzzy set theory adds gradation to this Boolean notion of set belonging. Thus it is the rare exception that a mathematical (optimization) model is a precise representation of the underlying system even if the symbols represent real numbers or real valued distributions. It is rare that an optimal solution from an optimization model is in reality best. Herbert Simon in many places (in particular see [46], p. 35, 36) states,

> "Of course the decision that is optimal in the simplified model will seldom be optimal in the real world. The decision maker has a choice between optimal decisions for an imaginary simplified world or decisions that are 'good enough,' that *satisfice,* for a world approximating the complex real one more closely. ... What a person *cannot* do he *will not* do, no matter how much he wants to do it. Normative economics has shown that exact solutions to the larger optimization problems of the real world are simply not within reach or sight. In the face of this complexity the real-world business firm turns to procedures that find good enough answer to questions whose best answers are unknowable. Thus normative microeconomics, by showing real-world optimization to be impossible, demonstrates that economic man is in fact a satisficer, a person who accepts 'good enough' alternatives, not because he prefers less to more but because he has no choice."

From an email discussion, Rommelfanger [43] relates the following.

> "In fact Herbert Simon develops a decision making approach which he calls the *Concept of Bounded Rationality.* He formulated the following two theses. **Thesis 1**: In general a human being does not strive for optimal decisions, but he tends to choose a course of action that meets minimum standards for satisfaction. The reason for this is that truly rational research can never be completed. **Thesis 2**: Courses of alternative actions and consequences are in general not known *a priori*, but they must be found by means of a search procedure."

The central tenet of this book is that most optimization is and should be a *satisficing* process. To this end, fuzzy and possibilistic optimization play a key, if not the most important, role. Why is fuzzy/possibilistic optimization not an explicit part of every Operations Research and Optimization curriculum? Why is it not in wide-spread practice yet? Is it that the practitioners do not know these new theoretical instruments? Is it that too many university professors do not understand the new concepts? Is it that clear and compelling "industrial strength" models have yet to appear in a way to make an impact? There are some "industrial strength" models (see the application section of this volume and [31], [38]). This volume presents the compelling reasons for including fuzzy and possibilistic optimization at the heart of normative decision theory.

We first define two terms that are frequently used in conjunction with fuzzy set theory and possibility theory, "uncertainty" and "incomplete information." From [14], we have:

Definition 1. *A piece of information is said to be **incomplete** in a given context if it is not sufficient to allow the agent to answer a relevant question in this context. A piece of information is **uncertain** for an agent when the latter does not know whether this piece of information is true or false.*

2 Fuzzy Set Theory and Possibility Theory in Optimization

This section begins with an example "industrial strength" application which shows the applicability and relevance of fuzzy and possibilistic optimization in an actual application. This problem was reported in [31], [38].

Example 2. *(Radiation Therapy of Tumors - [31]) The determination of how to use particle beams to treat tumors is called the radiation therapy problem (RTP). Beams of particles, usually photons or electrons, are oriented at various angles and with varying intensities to deposit dose (energy/unit mass) to the tumor. The idea is to deposit just enough radiation to the tumor to kill all the tumor cells while sparing normal tissue. The process begins with the patient's computed tomography (CT) scan. Each CT image is examined to identify and contour the tumor and normal structures. The image subsequently is vectorized. Likewise, candidate beams are discretized into beamlets, where each beamlet is the width of a CT pixel. A pixel is the mathematical entity or structure (a square in the two-dimensional case and a cube in three dimensions) that is used to represent a unit area or volume of the body at a particular location. For two-dimensional problems, about seventeen CT scans (slices) are sequentially "stacked" (to form a three-dimensional image that covers the tumor), and a variety of resolutions might be considered, $64 \times 64, 128 \times 128, 256 \times 256, 512 \times 512$. One set of beams each at ten or more equally spaced angles is not unusual. Since we constrain the dosage at*

each pixel, for ten equally spaced angles, the complexity of the problem ranges from an order of $17 \cdot 10 \cdot 64^2$ to $17 \cdot 10 \cdot 512^2$ potential constraints. However, since all pixels are not in the paths of the radiation beams that hit the tumor, and some are outside the body, we a priori set the delivered dosages at these pixels to zero and remove them from our analysis. This corresponds to blocking the beam, which is always done in practice. The identification of a set of beam angles and weights that provide a lethal dose to the tumor cells, while sparing healthy tissue with a resulting dose distribution acceptable and approved by the attendant radiation oncologist, is called a treatment plan. The largest actual problem solved is on the order of $500,000$ constraints in a fuzzy/possibilistic optimization problem. A dose transfer matrix A^T (representing how one unit of radiation intensity in each beamlet is deposited in pixels - for historical reasons, we use a transpose to emphasize its origin as the discrete version of the inverse Radon transform), called here the attenuation matrix, specific to the patient's geometry, is formed where columns of A^T correspond to the beamlets and rows represent pixels. A component of a column of the matrix A^T is non-zero if the corresponding beamlet intersects a pixel, in which case the value is the positive fraction of the area of the intersection of the pixels with the beamlet otherwise it is zero. The beams then are attenuated according to a factor dependent on the distance from where the beam enters the body to a pixel within the body and the type of tissue at that pixel. The variables are vectors x that represent the beamlet intensities. The mathematical programming problem is

$$z = \min f(c, x)$$
$$A^T x \leq b.$$

For this problem, there are four places in which the data turn this problem into a fuzzy/possibilistic optimization problem. First, the objective function can be a probability function (upper/lower bounding), the probability that the radiation intensity vector x will turn a health pixel into a cancerous one. Each row of the left side, that is, each row of the constraint matrix A^T, represents each pixel in the path of the beam, the "beam's eye view" of the tumor. Thus, each row accumulates pixel by pixel total radiation deposited by the radiation intensity vector x at that pixel. This will occur mathematically since the ith row vector (the ith pixel) A_i^T when dotted into the vector x is the sum of radiation at the pixel, $A_i^T x$. Since a pixel can be cancerous, or cancerous to a degree (the boundaries between cancerous and non-cancerous are gradual, transitional and thus fuzzy), the left side matrix A^T matrix is composed of fuzzy intervals. The right-hand side value is the maximum allowable dosage (for critical organs in the path of the beam and maximal value representing pre-burning for the tumor cells). Separately, for tumor cells, there is also an associated minimal value, the smallest value a radiation oncologist does not allow the radiation to go below (it is the minimal acceptable radiation that will kill a cancer cell). These values may be considered to be possibilistic

since these are values derived from research, expert knowledge (epistemic values), and experience with preferred values within the range. Therefore, the right-hand side value may be considered either as a target (with preferred ranges) or as purely as a possibilistic value.

A radiation oncologist is a *satisficer.* Any good treatment regime (identified by the mathematical programming problem) will suffice. Of course the optimal would be better, but given the variations in the data and model, "best" is an illusive value and arguably unattainable. The above example problem illustrates fuzzy, probability, and possibility optimization in an applied problem which has been solved (see [31], [38]).

Possibility theory as a mathematical theory is based on a set of axioms or properties found below. It is used to model systems of entities or variables that are uncertain due to *knowledge deficit* (inherent or acquired) or *incomplete information.* Fuzzy set theory deals with sets which are a generalization of the classical notion of set. It is used to model systems of entities or variables whose belonging (to a set) is *gradual* or *transitional.* Underlying possibility theory is the principle of minimal specificity (see [16]). That is, when the values of parameters or variables are not completely specified (for whatever reason - by choice, by finances, by lack of the ability to obtain the precise value, by the fact that inherently a precise value cannot be obtained), is there still sufficient structure to the information that is available for mathematical analysis? One of these mathematical structures is possibility theory.

The following simple example shows that *probability alone* is insufficient to describe uncertainty of every type. Suppose all that is known is that

$$x \in [1, 4]. \tag{1}$$

Clearly, $x \in [1, 4]$ implies that the real value that x represents is not certain (albeit bounded). If the uncertainty that $x \in [1, 4]$ represents were probabilistic (x is a random variable that lies in this interval), then every distribution having support contained in $[1, 4]$ would be equally valid given (1). Thus, if one chooses the uniform probability density distribution on $[1, 4]$,

$$p(x) = \begin{cases} \frac{1}{3} & 1 \leq x \leq 4 \\ 0 & otherwise, \end{cases}$$

which is the usual choice given no other indication other than the support, one gives up information. The approach that keeps the entire uncertainty of (1) considers it as all distributions whose support is $[1, 4]$. The pair of cumulative distributions that bound **all** cumulative distributions with this given support is depicted in Figure 1. The statement $x \in [1, 4]$ not only represents a random variable whose support is $[1, 4]$, but it can be a mathematical entity, an interval. When $[x] = [1, 4]$ is a mathematical entity, an interval, the statement $x \in [1, 4]$ has no associated uncertainty. It is complete, precise, and coherent in contradistinction with $[1, 4]$ containing all probability density

functions whose support is contained in this interval. It is true that the same object, $[1, 4]$, interpreted probabilistically and as an interval, has two semantically distinct and analytically distinct meanings, they have a different calculus, metric, convergence structure, and so on.
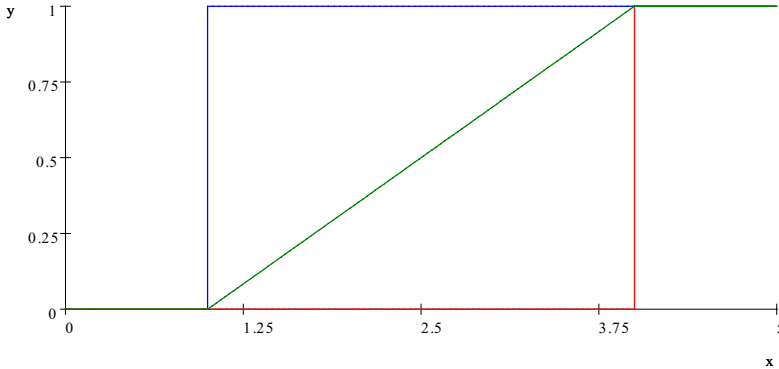

Fig. 1 BOUNDING CUMULATIVE DISTRIBUTIONS - Possibility (blue), Necessity (red), Uniform (green)

The upper cumulative distribution depicted in Figure 1 is a *possibility distribution* (blue), and the lower cumulative distribution is a *necessity distribution* (red). When the statement $x \in [1, 4]$ represents an unknown random variable whose support is the interval, to keep all the information about the uncertainty not only requires a pair of bounding functions, but a different arithmetic and mathematical analysis than "simply" functional analysis on probability distributions. The uniform distribution is precisely the intuitive solution to lack of information, "Choose the midpoint of the distribution pair as the solution *if one has to choose*." Of course, the case is made here that $x \in [1, 4]$ can also be tied to uncertainty which is purely non-probabilistic information deficiency in addition to an uncountably infinite set of random variables (whose support is contained in this interval).

2.1 Fuzzy Set Theory

Fuzzy sets are sets in which the Boolean property of belonging that characterizes classical sets is generalized to allow degrees of belonging continuously from zero (indicating not belonging for sure) to one (indicating belonging for sure). A classical set is one in which every element in the universal set has a degree of belonging that is described by the characteristic function (zero/one), whereas a fuzzy set is one in which every element in the universal set is described by a function, called a *membership function*, whose range is

between zero and one. A fuzzy set $A \in \Omega$ (universal set) is uniquely described by its membership function

$$\mu_A(x) : \Omega \to [0,1], x \in \Omega. \tag{2}$$

A classical set is one whose membership function is,

$$\mu_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases} \tag{3}$$

The general definition of a fuzzy set (2) does not impose any condition except that the fuzzy membership be a function. In the context of optimization, we will restrict ourselves to fuzzy numbers and fuzzy intervals, and the membership function must be upper/lower semi-continuous (or more practically continuous), where at least one value of the domain has a membership value of one. In contrast to probability for finite spaces (see discussion in [16]), for an event $x \in A$, "... $prob(A)$ is the probability that an ill-known single-valued variable x ranging in Ω hits the fixed well-defined set A." For example, suppose we are rolling two die, and $A = \{4, 5\}$. The probability $prob(A) = \frac{7}{36}$. Here x, which ranges in $\Omega = \{2, 3, ..., 12\}$, is unknown (it is the outcome of a throw of the dice), but the set A is precisely known. When we consider $\mu_A(x)$, then x is "fixed", known precisely, whereas the set A is ill-defined (transitional).

[8] A fuzzy set (2) is an abstract mathematical notion. It presupposes nothing about what it could be applied to the notion. It corresponds to the ideal of non-Boolean predicates. The idea of gradualness or grade (non-Boolean) is opposed to "all or nothing" (Boolean). Thus, fuzzy means *gradual* and not vague or uncertain. So, consider a set. Some sets are real entities, for example, the set of all salaried tenured professors who are employed by the Mathematical and Statistical Sciences Department at the University Colorado Denver during the 2008-2009 academic year or the set of *older* mathematicians. Some of these real sets are fuzzy, for example, the set of older mathematicians, since the concept of "older" is gradual, graded. These (fuzzy) sets are taken as "lumped" entities (*older* is a "lumped" entity). They are a *conjunctive or linked* set of elements having more or less weight.

Let us look again at our initial example. The interval $[1, 4]$ considered as a fuzzy set has membership function

$$\mu_{[1,4]}(x) = \begin{cases} 1 & x \in [1, 4] \\ 0 & otherwise. \end{cases}$$

Considered as a fuzzy set, the elements of $[1, 4]$ are "lumped" or linked into one entity, an interval. This intertwined or conjunctive property characterizing fuzzy sets is contrasted with the "mutually exclusive", unlinked, or

disjunctive property that characterizes possibility theory and probability theory, as we shall discuss further below.

Mathematical analysis on fuzzy sets (here we are thinking of fuzzy sets as abstract mathematical entities, that is, sets of membership functions (2)) require a measure, a fuzzy measure, which is used to define order, extent, convergence, integrals (transformation of a function into a number), and so on. Fuzzy measures and integrals are used in fuzzy logic to compare/order and to "defuzzify" (turn a fuzzy set into a number and thus an action - "shift into overdrive" for example). This is akin to determining the expected value for continuous distributions in probability theory that transforms a (probability distribution) function into a single real number (the mean).

Definition 3. *[30] Given a universal set Ω and a nonempty family $\mathcal{F}$ of subsets of Ω, a **fuzzy measure** on $\langle \Omega, \mathcal{F} \rangle$ is a function*

$$g : \mathcal{F} \to [0, 1]$$

that satisfies the following properties:
1) $g(\varnothing) = 0$, and $g(\Omega) = 1$ (boundary conditions)
2) $\forall A, B \in \mathcal{F}$, if $A \subset B$, then $g(A) \leq g(B)$ (monotonicity)
3) $\forall$ increasing sequence $A_1 \subset A_2 \subset ...$ in $\mathcal{F}$, if $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$, then

$$\lim_{i \to \infty} g(A_i) = g\left(\bigcup_{i=1}^{\infty} A_i \right) \text{ (continuity from below)}$$

4) $\forall$ decreasing sequence $A_1 \supset A_2 \supset ...$ in $\mathcal{F}$, if $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$, then

$$\lim_{i \to \infty} g(A_i) = g\left(\bigcap_{i=1}^{\infty} A_i \right) \text{ (continuity from above).}$$

In the above, 3) and 4) can be viewed as endowing consistency to fuzzy measures. Usually, the fuzzy measures are defined over a structured set of sets such as a ring or a σ-algebra or the full power set $P(\Omega)$. Fuzzy measures may be considered as generalizations of probability measures and, in the broader sense, classical measures. The difference between the fuzzy measure and a probability measure is that the additivity condition is replaced by a weaker one of monotonicity and continuity.

Property 2) of Definition 3 implies that $g(A \cup B) \geq \max\{g(A), g(B)\}$ and $g(A \cap B) \leq \min\{g(A), g(B)\}$. A fuzzy measure is a real valued measure that is used for mathematical analysis of fuzzy sets. The semantic associated with fuzzy measure is that the assigned number $g(A)$ to the set $A \in \mathcal{F} \subset \Omega$ indicates the degree to which the given element, set A of $\mathcal{F}$, whose characterization is transitional, belongs to $\mathcal{F}$.

2.2 Possibility Theory

The mathematical models of systems which are described by the knowledge we as humans have (epistemic) about the system are often not probabilistic. Deterministic models are (supposed to be) precise statements about properties of the real system, the thing itself. They refer to the real world. Possibilistic models often refer to the knowledge humans have about the system. The model exists or refers to what is in someone's (or a group of people's) head, not the actual system. Possibility is a mathematical structure with a set of operations that allow us to model some uncertainties that are not probabilistic, uncertainties that are not random, frequencies, or chance. However, as in the interval example, they also provide a structure for probabilistic bounds which themselves are derived from (cumulative) probabilities. Even when the statement of the problem is clearly probabilistic (the probability that radiation intensity x will turn a health cell into a cancerous cell), a single probability distribution function is impossible to obtain for every cell, and for every human body, and for every bodily condition (obesity, anorexia, lean, and so on), and for every type of cancer, and every type of radiation type (electron, proton, neutron), and must be approximated. To obtain one single probability distribution function $p(x)$ is an approximation at best. It is perhaps more useful to find bounding functions and to do our analysis on these rather than committing to a single probability at the start.

Some set representations are epistemic entities, the information possessed by people about real entities. These set representations do not refer to the real world, but to what people know about the real world. Dubois uses the following example. I may "know" epistemically that the Prime Minister of France is between 40 and 55 years of age. This is the extent of my knowledge of the age of the prime minister of France, which may be wrong, and it may differ from your knowledge. The interval [40,55] refers to an interval of real ages. However, the age of Prime Minister Sarkozy, which exists, is a *point,* not a set. The set [40, 55] is the set of possible ages of Prime Minister Sarkozy according to my knowledge, where the actual age of the Prime minister is a point. My interval of possible ages in this case is not a conjunction of elements, it is a disjunction of elements (the set of distinct numbers in the interval). A generalized characteristic function of this set is a possibility distribution, despite the fact that there is only one age of Prime Minister Sarkozy. If some values of ages in the interval [40, 55] are more plausible than others, then this plausibility defines a preferential function on [40, 55] which is a possibility distribution.

Possibility is also an abstract mathematical structure which is independent of applications.

Definition 4. *[55] Let* **P(Ω)** *denote the power set of the universal set Ω (although a σ-algebra would also work). A **possibility measure** Pos:***P(Ω)** → *[0,1] has the following properties:*

1) $Pos(\varnothing)=0$, $Pos(\Omega)=1$

2) $A \subset B \Rightarrow Pos(A) \leq Pos(B)$, for $A, B \in \Omega$.

3) $Pos(\bigcup_{i \in I} A_i) = \sup_{i \in I} Pos(A_i)$.

Possibility is a non-additive measure like fuzzy measures, but it differs from a fuzzy measure since the two continuity conditions are replaced by the supremum condition on the union. Moreover, property 3) of Definition 4 implies that

$$Pos(A \cup B) = \max\{Pos(A), Pos(B)\}.$$

As an abstract structure, it also differs from a probability measure because the additivity condition is replaced by the supremum condition. Besides defining possibility via the three properties given above, possibility measures also can be constructed beginning from probability in four different ways one of which as was given by our first example [1, 4]. The other three ways are presented in the sequel.

Any possibility measure is determined uniquely by a given possibility (distribution) function

$$\pi : \Omega \to [0, 1]$$

via the formula

$$Pos(A) = \sup_{x \in A} \pi(x), A \subset \Omega. \tag{4}$$

Also, given a possibility measure, we can define a possibility (distribution) function $\pi : \Omega \to [0, 1]$

$$\pi(x) = Pos(\{x\}). \tag{5}$$

That is, we can go from a measure to a possibility distribution function in a natural way. Moreover, possibility measures are distinct from fuzzy measures (see [41]).

It turns out that Dempster-Shafer's plausibility and belief functions (see [6] and [45]) using an auxiliary probability assignment function with the requirement that the focal elements (elements whose probability assignment functions are non-zero) be nested. When this occurs, the resultant is a possibility measure. Thus, possibility theory can be considered as a particular case of Dempster-Shafer theory. To see this, consider a finite universal set Ω, for the sake of simplicity, and where $P(\Omega)$ denotes the power set of Ω. We restate possibility and necessity in this context (finite universe) and call a function $Nec : P(\Omega) \to [0, 1]$ a *necessity measure* if and only if

$$Nec(\emptyset) = 0$$
$$Nec(\Omega) = 1$$
$$Nec(A \cap B) = \min\{Nec(A), Nec(B)\}, \tag{6}$$

for all $A, B \in P(\Omega)$. A function $Pos : P(\Omega) \to [0, 1]$ is called a *possibility measure* if and only if,

$$Pos(\emptyset) = 0$$
$$Pos(\Omega) = 1$$
$$Pos(A \cup B) = \max\{Pos(A), Pos(B)\}, \tag{7}$$

for all $A, B \in P(\Omega)$. Note that (7) implies that if $A \subseteq B, Pos(A) \leq Pos(B)$. Moreover,

$$Nec(A) = 1 - Pos(A^c).$$

The relationship between necessity and Dempster-Shafer theory is as follows. Evidence theory is based on two non-additive measures on a universal set Ω, a belief measure, Bel, and a plausibility measure, Pl, defined as (see [30])

$$Bel : P(\Omega) \to [0, 1] \tag{8}$$
$$Bel(\emptyset) = 0, Bel(\Omega) = 1,$$
$$Bel(\bigcup_{i=1}^{n} A_i) \geq \sum_i Bel(A_i) - \sum_{i<j} Bel(A_i \cap A_j) + \ldots$$
$$+ (-1)^{n+1} Bel(\bigcap_{i=1}^{n} A_i)$$

for all possible families of subsets of Ω. Belief is superadditive. The dual of belief is plausibility which is defined

$$Pl(A) = 1 - Bel(A^C). \tag{9}$$

A function $m : P(\Omega) \to [0, 1]$ is called a *probability assignment function* if and only if

$$m(\emptyset) = 0 \tag{10}$$
$$\sum_{A \in P(\Omega)} m(A) = 1.$$

Lemma 5. *[45] Given a probability assignment function, a Belief and Plausibility measure on Ω can be defined*

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{11}$$

for all $A \in P(\Omega)$, and

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B). \tag{12}$$

Conversely, given a belief measure, a probability assignment function can be defined by

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} Bel(B), \qquad (13)$$

where $|A - B|$ denotes the cardinality of $A - B$, for all $A, B \in P(\Omega)$.

Essentially, starting from probability, we start from (10) where the probabilities are known on *sets* of the space, **not** elements of the space, with no particular assumed structure (they may be overlapping, for example) other than that the sums of their probabilities add to one. From this partial information, we can derive a mathematical structure which does not, in general, have the additive property of probabilities which is called Belief and Plausibility. On the other hand, we can start from the point of view that we have a sub/super additive set functions called Belief and Plausibility (whose properties are is defined by (8) and (9)) and derive a probability assignment function from (13). If the assignment function is known on *elements* of the space, then belief is equal to plausibility, and the resulting structure is probability (with additivity).

The subsets of Ω with positive values of a given probability assignment function are called *focal elements*. Suppose all focal elements of the basic assignment functions (13) can be ordered so that they are nested. That is, suppose we have an order to the focal elements determined by (13) so that the finite sequence

$$\{A_1, A_2, ..., A_K\}$$

and $m(A_i) > 0$ has the property that $A_i \subseteq A_j$ for $i \leq j$. Then

$$Nec(A) = \sum_{B \subseteq A} m(B) \qquad (14)$$

is a necessity measure. Moreover,

$$Pos(A) = \sum_{A \cap B \neq \emptyset} m(B) \qquad (15)$$

is a possibility measure. In other words, (10) can be given and *Bel* and *Pl* constructed via (11) and (12). Or *Bel* (and *Pl*) can be given and the probability assignment function be constructed via (13). Moreover, when we have nested sets, we take the focal elements of the probability assignment function on these nested sets and construct the possibility and necessity which are precisely the belief and plausibility measures.

We can obtain a probability assignment function from a given necessity measure by using (13) so that

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} Nec(B).$$

Thus, necessity measures are belief measures when the focal elements are nested. Fuzzy intervals are the collection of nested focal elements constructed from the non-zero $\alpha - cuts$. That is, for $\alpha > 0$, in the discretized case, let

$$0 = \alpha_0 < \alpha_1 < ... < \alpha_j < \alpha_{j+1} < ... < \alpha_N = 1,$$
$$m(A_{\alpha_j}) = \alpha_j - \alpha_{j-1}, j = 1, ..., N$$
$$A_\alpha = \{x | \mu_{fuzint}(x) \geq \alpha\},$$

where μ is the membership function of the fuzzy interval. Hence, possibility theory may be viewed as a special type of evidence theory. This allows possibility theory to be considered from a generalized probabilistic point of reference. Zadeh [55] defined only possibility measures/distributions. Dubois and Prade [11] were the first to develop the dual to possibility, necessity.

The previous section outlined several ways that lead to possibility distributions. There is a development (see [4]) which takes the possibility measure given above and develops the full mathematical structure such that the mathematical analysis is well-defined. We are thinking of possibility theory applied to mathematical analysis (optimization in particular). To this end, we construct possibility and necessity distributions in one of the following ways:

1. [37] Given a set of probabilities $\Omega = \{p_\alpha(x), x \in \mathbb{R}, \alpha \in I$, where I is an index set$\}$,

$$Pos(x) = \sup_{\alpha \in I} p_\alpha(x)$$
$$Nec(x) = \inf_{\alpha \in I} p_\alpha(x).$$

2. [28] Given an unknown probability $p(x)$ which is known to exist inside a bounding pair of functions $p(x) \in [\underline{f}(x), \overline{f}(x)]$, construct necessity/possibility distributions such that $p(x) \in [Nec(x), Pos(x)]$.
3. [6], [45] Given a probability assignment function m whose focal elements are nested, construct necessity/possibility distributions according (15) and (14).
4. A fuzzy interval, defined below, generates a possibility and necessity pair. Fuzzy intervals, as used here, are what is called in most of the literature fuzzy numbers. The possibility and necessity functions are constructed as was done from our initial example [1, 4] (also see Figure 3 below).

The most prevalent approach is to define the entities of interest in optimization (the coefficients and/or the right-hand side values, for example) to be fuzzy intervals in which case they will be able to capture both gradualness or transition and lack of specificity/information as we specify below. Thus, possibility distributions used in possibilistic optimization typically are associated with the membership function of fuzzy numbers, whose generalization is called a *fuzzy interval*. If the coefficients arise from probability-based

possibility (as in items 1-3 listed above), then this generates upper and lower possibilistic optimization (see [37]). A fuzzy number is a fuzzy set with upper/lower semi-continuous membership function with one and only one value, x^*, such that $\mu(x^*) = 1$ (where x^* is the "fuzzified" number). The set of numbers for which the membership value are one is called the *core*. So for example, a fuzzy number 2 would have $\mu(2) = 1$, which is depicted in Figure 2.
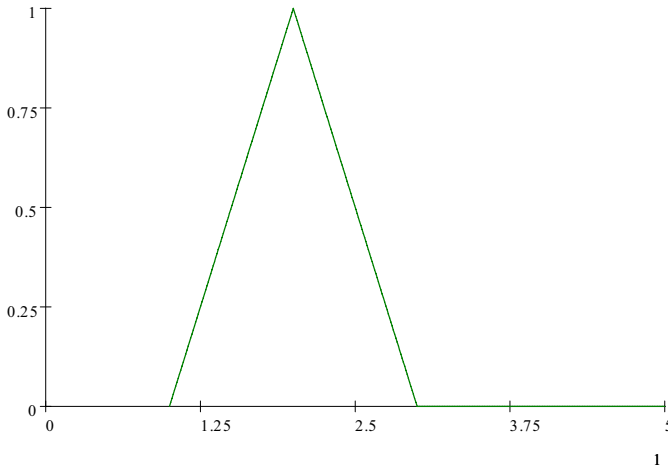

Fig. 2 A Fuzzy number two

A fuzzy interval M, depicted as a trapezoid in Figure 3, is a fuzzy number except the core (which must also exist) does not have to be a singleton. There are various views (applications) of fuzzy intervals. A fuzzy interval can be used to enclose a set of probability distributions where the bounds are constructed from the fuzzy interval (blue line being the possibility and green line being the necessity in Figure 3). The core of the fuzzy interval is the top, the horizontal line segment between 2 and 3 at height 1, indicated in Figure 3. The possibility and necessity as indicated below enclose all probabilities whose upper limit is the possibility (blue line) and lower limit is the necessity (green line). Thus, according to [15],

"A fuzzy interval M can thus be viewed as encoding a family of probabilities, a set of probability measures P_M defined by

$$P_M = \{P | \Pi_M(A) = \sup_{a \in A} M(a) \geq prob(A), A \text{ measurable}\}.$$

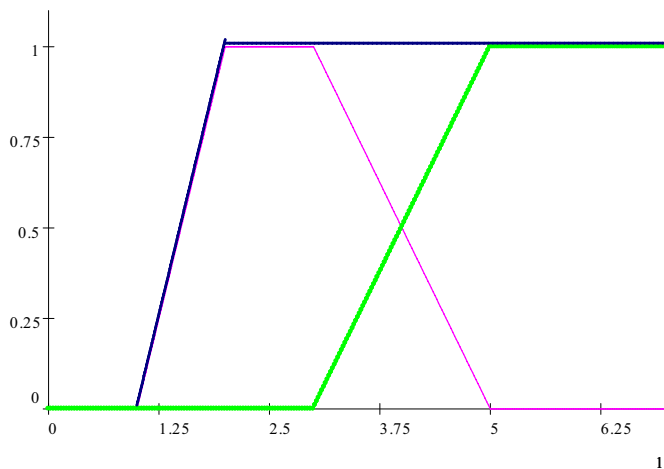It is important to notice that there are actually three probabilistic views of a fuzzy interval:

Fig. 3 Fuzzy interval - Megenta

a) The *imprecise probability* view whereby M encodes a set of probability measures shown in Figure 3.
b) The *pair of PDFs* view whereby M is defined by two random variables x^- and x^+ with cumulative distributions in blue and green of Figure 3 and M stands for the random interval $[x^-, x^+]$.
c) The *random set* view whereby M encodes the one point coverage function of a random interval, defined by the probability measure on the unit interval (for instance the uniformly distributed one) and a family of nested intervals (the $\alpha - cuts$), via a multivalued mapping from $(0,1]$ to $\mathbb{R}$, following Dempster [6]."

A reason that one might want to use probability-based possibility (interpretations a), b) or c)) rather than probability is precisely in situations for which real values or complete probability density functions for data are not available. For example: (1) we don't know which of a given set of probabilities to use, (2) all we know is that the probability is bounded by two functions, or (3) we do not have the probability distribution on singletons, but on sets. Whether an entity of interest inherently lacks specificity (the minimal radiation that will kill a particular patient's prostate tumor cell located at (x, y, z) is $Pos(x, y, z)$), lacks sufficient research to determine its precise value or its precise probability density function, its deterministic functional representation is not required, in the sense that one can get by with a more general form than its deterministic equivalent - perfect information, for the use to which it is put (the light wave reflection measured by a satellite sensor to impute the depth of the ocean, low/medium/high might suffice), or complexity reduction (low, medium, high speed for the automatic gear shifting mechanism on a car), lack of information/specificity is a part of many if not most

problems. Moreover, when we have models that are built from epistemic knowledge (human ideas about the system rather than the system itself), many of these linguistically derived models are possibilistic either in their entirety or partially.

2.3 Fuzzy Set Theory and Possibility Theory - Distinctions

The semantics of fuzzy sets and possibility are different. The difference between fuzzy set theory and possibility theory is that the semantics of fuzzy are tied to gradualness whereas the semantics of possibility refers to incomplete (deficient) information about an existent entity. As we have mentioned above, fuzzy is a non-Boolean set whose elements transitionally belong to a given set. Possibility is tied to incomplete information. A fuzzy set membership function uniquely describes the set and defines the degree to which an element belongs to the set. The possibility distribution assigns the degree to which the evidence supports the element's belonging to the set in question. In particular, consider the two statements

$$\mu_A(x) = 1,$$

and

$$\pi_A(x) = 1.$$

In the first instance, the membership function value of x is 1. In the second, the possibility distribution at x is 1. In the first case, it is certain that x belongs to the set A. In the second case, all that can be said is that the best information at hand indicates that it is most plausible (or possible) that x belongs to A. $\mu_A(x) = 0$ means that $x \notin A$ for sure. $\pi_A(x) = 0$ means that all the given evidence at hand indicates that x does not belong to A (or x is not A).

> [16] "Limited specificity can be modelled in a natural way by possibility theory. The mathematical structure of possibility theory equips fuzzy sets with set functions, conditioning tools, notions of independence/dependence, decision-making capabilities (lattices). Lack of information or lack of specificity means we do not have 'the negation of a proposition is improbable if and only if the proposition is probable.' In the setting of lack of specificity, 'the negation of a proposition is impossible if and only if the proposition is necessarily true.' Hence, in possibility theory pairs of possibility and necessity are used to capture the notions of plausibility and certainty. When pairs of functions are used we may be able to capture or model lack of information. *A membership function is a possibility only when the domain of a fuzzy set is decomposable into mutually exclusive elements.* A second difference (between probability and possibility) lies in the underlying assumption

regarding a probability distribution. Namely all values of positive probability are mutually exclusive. A fuzzy set is a conjunction of elements. For instance, in image processing, imprecise regions are often modelled by fuzzy sets. However, the pixels in the region are not mutually exclusive (although they do not overlap). Namely the region contains several pixels, not a single unknown one. *When the assumption of mutual exclusion of elements of a fuzzy set is explicitly made, then, and only then, the membership function is interpreted as a possibility distribution; this is the case of fuzzy numbers describing the ill-located unique value of a parameter.*" (my *italicized* emphases)

More recently, [14] state

"A set used for representing a piece of incomplete information is called a *disjunctive set.* It contrasts with a conjunctive view of a set considered as a collection of elements. A conjunctive set represents a precise piece of information. For instance, consider the quantity $v = sisters(Pierre)$ whose range is the set of subsets of possible names for Pierre's sisters. The piece of information $\{Marie, Sylvie\}$ is precise and means that Pierre's sisters are Marie and Sylvie. Indeed, the frame is then $S = 2^{NAMES}$, where $NAMES$ is the set of all female first names. In this setting, a piece of incomplete information would be encoded as a disjunction of conjunctive subsets of $NAMES$."

For example, an image could be segmented/classified into two sets, stomach lining and stomach muscle. Every pixel in the image is given a value v where $0 \leq v \leq 1$ with respect to being stomach lining or stomach muscle. This is conjunctive and thus a fuzzy set. That is, each pixel is stomach lining to specified degree (between 0 and 1), and (conjunction) each pixel is also stomach muscle to a specified degree (between 0 and 1). On the other hand, suppose we use the fuzzy trapezoid interval 59/59.9/60.1/61 to model the possibilistic notion of a tumorcidal dose to each tumor pixel. This is an incomplete set of information about each tumor pixel. That is, a tumor pixel has an associated *distribution*, a fuzzy interval, a function. A tumor pixel is a fuzzy interval 59/59.9/60.1/61. The pixel is not represented by a number (as in the case of stomach lining or muscle - there are two numbers one for each class), the pixel is represented by an entire distribution. Moreover, possibility is always normalized since the semantics of possibility is tied to an existential entity. Thus, *not all fuzzy set membership functions can be transformed into possibility distributions.*

2.4 Fuzzy Set Theory and Possibility Theory in Optimization

Gradualness characterizes many linguistic descriptions of what we know about properties of entities and systems of these entities. Indeed, classification of

reality, which often is the first step in mathematical modelling, may reflect the inherent transitional nature of the class as well as objects in that class. Thus, "tumor" is a classification which is "intertwined." Each "tumor" pixel within a CT scan may consist of fully cancerous cells in a pixel and (conjunctive) pre-cancerous cells coexisting in the same pixel.

Mathematical modelling has two directions - one toward more specificity (a real value being one "end point" of specificity) or more generality (the universe being at the other end of specificity). Generality is often useful to simplify analysis and reduce the complexity. The greatest use and application of fuzzy set theory is fuzzy logic controllers. In mathematical analysis, it can be argued that fuzzy optimization has been the most successful application. Fuzzy optimization is an extension to flexible programming and allows for a broader and more ample approach to flexible programming.

The appropriate classification of *possibilistic optimization* is *optimization under uncertainty,* where some (or all) the input data (parameters) to the optimization model lack specificity, and/or the information is insufficient to yield a real valued number or a probability distribution. This is distinguished from *fuzzy optimization* that appropriately belongs in the class of *flexible programming* problems. The uncertain parameters that are possibilistic must adhere to (7), which means that more information about the parameter can never yield less certainty (if $A \subseteq B, P(A) \leq P(B)$), and these sets must be nested. This property (more information leading to greater certainty) is not axiomatically present in fuzzy measures. For example, the fuzzy set "older" remains transitionally "older," regardless of how much more information about older is obtained. However, the interval of the age of the prime minister of France will narrow with more (correct) information.

Not all systems possess this property of more information not increasing uncertainty. In competitive markets, (cold) wars, adversarial relationships (such as political campaigns, trials, or propaganda), more information might be worse given the propensity to disinformation, deceptive information, or lying all mixed together with legitimate information. When the parameters are defined as fuzzy intervals, they are also possibility distributions and automatically adhere to the axioms (definition) of possibility even if they are fuzzy entities. When fuzzy intervals are used for the parameters, what must be checked in terms of appropriate use is the semantics. The semantics associated with possibilistic uncertainty must be tied to information deficiency (lack of specificity).

The next set of paragraphs are taken from [35] both directly and in modified form. By fuzzy and possibilistic optimization, as used here, we mean optimization when at least one element of the input data is a real valued interval, a real valued random variable, a real valued fuzzy number, or a real valued number described by a possibility/necessity distribution. The use of necessity distributions are done similarly to possibility except that the necessity sematic is a pessimistic one, while the possibility semantic is optimistic.

We consider the following general programming problem:

$$z = \min f(x, c)$$
$$\text{subject to} \quad g_i(x, a) \le b \quad i = 1, ..., M_1 \tag{16}$$
$$h_j(x, d) = e \quad j = 1, ..., M_2.$$

The constraint set is denoted $\Omega = \{x \mid g_i(x, a) \le b \ i = 1, ..., M_1, h_j(x, d) = e$ $j = 1, ..., M_2\}$. It is assumed that $\Omega \ne \emptyset$. The values of a, b, c, d, and e are inputs (data, coefficients, parameters) of the programming problem. These values are subject to uncertainty for a variety of reasons. Depending on the nature of the uncertainty, they may be probability distributions, intervals, fuzzy sets or possibilistic distributions. Moreover, the operator min and relationships $=$ and $\le$ can take on a flexible or fuzzy meaning becoming a soft relationship or constraint. For example, the equality and inequality relationships may be aspirations, that is, they may take on the meaning, "Come as close as possible to satisfying the relationships with some degree of violation being permissible." On the other hand, the value of $a, b, c, d,$ or e may be described by a probability, interval, fuzzy or possibilistic distribution. In either case, the meaning of the relationships must be specified. When the objective function and/or constraints are defined by functions other than real valued convex functions, the optimization problem may not be (undoubted is not) convex so that typical solution methods are local. In very simple cases where the constraint is of the form $Ax \le b$, and the coefficients of the matrix and right-hand side values are intervals, the solution set can be a star-shaped region (see [20]). Recall that an interval is a fuzzy number. Moreover, when the components of the matrix A are other than real valued, this means that the underlying model as specified by linear relationships is not known exactly or that the model is precise, but knowledge of what the value of the data are incomplete. We use a tilde, ˜, to denote a fuzzy set, and a "hat", ˆ, to denote a possibility distribution.

2.4.1 Fuzzy and Possibilistic Optimization Semantics

Next what is meant by decision-making in the presence of fuzzy and possibilistic entities is defined. These definitions are central to the semantics and methods. In their book (Chapter 5) Dubois and Prade [9] give clear definitions and distinctions of fuzzy measures, possibility and probabilities often forgotten and ignored by researchers (see also Chapters 1 and 7 of [13]).

1. *Fuzzy Decision Making:* Given the set of real valued (crisp) decisions, Ω, and fuzzy sets, $\{\tilde{F}_i \mid i = 1 \text{ to } n\}$, find the optimal decision in the set Ω. That is,

$$\sup_{x \in \Omega} h\left(\tilde{F}_1(x), ..., \tilde{F}_n(x)\right), \tag{17}$$

where $h : [0, 1]^n \to [0, 1]$ is an aggregation operator [30], often taken to be the *min* operator, and $\tilde{F}_i(x) \in [0, 1]$ is the fuzzy membership of x in

fuzzy set $\tilde{F}_i$. The decision space Ω is a set of real numbers (*crisp set*), and the optimal decision satisfies a mutual membership condition defined by the aggregation operator h. This is the method of Bellman and Zadeh [1], Tanaka, Okuda and Asai [48], [49], and Zimmermann [56], who were the first to develop fuzzy mathematical programming. While the aggregation operator h historically has been the *min* operator, it can be, for example, any $t - norm$ that is consistent with the context of the problem and/or decision methods (see [29]).

2. *Possibilistic Decision Making:* Given the set of real valued (crisp) decisions, Ω, and the set of possibility distributions representing the uncertain outcomes from selecting decision $\mathbf{x} = (x_1, ..., x_n)^T$ denoted $\Psi_x = \{\hat{F}_x^i, i = 1, \ldots, n\}$, find the optimal decision that produces the best set of possible outcomes with respect to an ordering U of the outcomes. That is,

$$\sup_{\Psi_x \in \Psi} U(\Psi_x), \tag{18}$$

where $U(\Psi_x)$ represents a "utility" of the set of distributions of possible outcomes $\Psi = \{\Psi_x | x \in \Omega\}$. The decision space Ψ is a **set of possibility distributions** $\Psi_x : \Omega \to [0, 1]$ resulting from taking decision $x \in \Omega$. This is the semantic taken in the possibilistic optimization of Inuiguchi [22], [23], [24] and Jamison and Lodwick [27]. If $\hat{F}_x = \hat{2}x_1 + \hat{3}x_2$, then each $\mathbf{x} = (x_1, x_2)^T$ generates the possibility distribution $\hat{F}_x = \hat{2}x_1 + \hat{3}x_2$.

Remark 6. *Let us summarize what we have just presented. For fuzzy sets $\tilde{F}_i$, $i = 1$, ..., n, given x, $[\tilde{F}_1(x), ..., \tilde{F}_n(x)]^T$ is a real valued **vector**. Thus, we need a way to aggregate the components of the vectors into a single real value. This is done by a t-norm, min for example. For possibility, given x, $\Psi_x = \{\hat{F}_x^i, i = 1, \ldots, n\}$ is a set of **distributions**, so we need a way to turn this set of distributions into a single real value. This is done by the utility function, a generalized expectation, for example.*

Very simply, fuzzy decision-making selects from a set of real valued, crisp, elements ordered by an aggregation operator on corresponding membership functions, while possibilistic decision making selects from a set of distributions measured by a utility operator that orders the corresponding distributions. These two different approaches have two different ordering operators (an aggregation operation for fuzzy sets such as *min* and a utility function in the case of possibility such as a generalized expectation) and lead to two different optimization methods (see [35]). The underlying sets associated with fuzzy decision-making are fuzzy, where one forms the decision space of real valued elements from operations ("*min*" and "*and*", for example, in the case of optimization of [1], [49] and [56]) on these fuzzy sets. The underlying sets associated with possibilistic decision making are real value sets, where one forms the decision space of (possibility) distributions from operations on real valued sets.

The construction of an appropriate utility function is a challenge. The axioms of utility theory as developed by Von Neumann and Morgenstern [52] are usually required. The type of utility function that is used is a challenge and decision maker dependent. For example, if one is radiating a tumor that is quite aggressive, one's utility might have higher risk (the first derivative is large and positive over the domain) than if one were radiating a tumor that was growing very slowly (the first derivative is small and positive over the domain). For this presentation, we put aside the question of how to obtain an appropriate utility function noting that it is a key to the successful implementation of the methods contained herein. The key point is that in possibilistic optimization, one is using a utility such as a generalization of the expectation to transform distributions into one real valued function (which is then optimized), whereas in fuzzy optimization, one is using an aggregation operator such as a *min* or *t-norm* to transform vectors into one real valued function (which is then optimized).

The idea of the use of utility for decision making under uncertainty problems is discussed in [17] who show how to use two qualitative counterparts to the expected utility criterion, one type of utility, U, that can be used in (18), to express uncertainty and preferences in decision making under uncertainty. Thus, what is called here *possibilistic decision making,* (18), is related to what [17] develop. However, optimization as articulated here are quantitative methods (the mapping $U : \Psi \rightarrow \mathbb{R}$), whereas the focus of [17] is more qualitative.

2.4.2   Fuzzy Decision Making Using Fuzzy Optimization

Fuzzy decision making using fuzzy optimization was first operationalized by Tanaka, Okuda, and Asai (see [48], [49]) and Zimmermann (see [56]). This approach, based on the landmark theoretical paper by Bellman and Zadeh [1], relaxes systems of inequalities $Ax \leq b$ to denote aspirations. The results are *soft constraints,* where the number b to the right of the soft inequality is a target such that, if the constraint is less than or equal to b, the membership value is one (the constraint is satisfied with certainty), and, if the constraint is greater than $b + d$, (for an *a priori* given $d > 0$), the membership is zero (the constraint is not satisfied with certainty). In between, the membership function is interpolated so that it is consistent with the definition of a fuzzy number membership function in the context of the problem. Linear interpolation was the original form (see [56]). This models a fuzzy meaning of inequality that is translated into a fuzzy membership function and is the source of our use of the designation of **flexible programming** for these classes of problems. The $\alpha - level$ represents the degree of feasibility of the constraints, consistent with the aspiration that the inequality be less than b but definitely not more than $b + d$. Thus, the objective (according to [56]) is to simultaneously satisfy all constraints at the highest possible level

of feasibility as measured by the $\alpha-levels$ of the membership functions (that is "and" all membership functions).

The approach of Tanaka, Okuda, and Asai (see [48], [49]) and Zimmermann (see [56]) deals with one way to minimize constraint violations. However, their operationalization is not always Pareto optimal [8]. Their approach must be iterated - fix the constraints at bounds and re-optimize. Their method falls within a goal satisfaction approach in optimization in which the highest degree of goal attainment is sought. They do this by minimizing the violation of the most stringent constraint. Thus, for example, this approach may guarantee that every constraint is satisfied to a 0.65 degree or more, and it may be the case that every constraint is satisfied to the 0.65 level. However, it may also be that if one of the constraints were relaxed to a 0.6 constraint violation level, all others may be satisfied at a 0.95 level. That is, this approach does not look at the aggregate satisfaction, only the most constraining one. It is minimizing the maximum constraint violation.

An aggregate goal attainment tries to maximize an overall *measure* of aggregate goal satisfaction. The *aggregate* sum of goal attainment focuses on maximizing the cumulative satisfaction of the goals. The surprise function (see [38], [40]) is one such measure for an aggregate set of (fuzzy) goals. In particular, when the right-hand side values are interpreted as goals rather than rigid constraints, the problem may be translated into one of optimizing the aggregate goal satisfaction. Thus, for soft constraints derived as,

$$\text{hard}\ \ y_i = (A\mathbf{x})_i \leq b_i \Rightarrow \text{soft}\ \ y_i = (A\mathbf{x})_i \leq \tilde{b}_i, \tag{19}$$

where the right-hand side values of the soft constraint are *fuzzy numbers,* the transformation into a set of aggregate goal satisfaction problem using the surprise function as the measure for the cumulative goal satisfaction is attained as follows. A (soft) fuzzy inequality (19) is translated into a fuzzy membership function, $\mu_i(x)$, which is the possibility $pos(\tilde{b}_i \geq x)$. Each membership function is translated into a surprise by

$$s_i(x) = (\frac{1}{\mu_i(x)} - 1)^2. \tag{20}$$

These functions are added to obtain a total surprise

$$S(\mathbf{x}) = \sum_i s_i(y_i) = \sum_i s_i(y_i) = \sum_i s_i((A\mathbf{x})_i). \tag{21}$$

Note that (21) is an aggregation operator. A best compromise solution based on the surprise function is given by the nonlinear optimization problem

$$\min\ z = \sum_i s_i((A\mathbf{x})_i)$$

subject to $x \in \Omega$ (possible hard constraints).

That is, a real valued inequality constraint whose right-hand side value is a fuzzy number is translated into a fuzzy set. This fuzzy set then is transformed into a real valued set, one for each $\alpha - level$, via a surprise function whose domain, of course, is the $\alpha - level$ of the fuzzy sets. The objective is to minimize the **sum** of all surprise function values. Unlike Tanaka and Zimmermann, the constraints are not restricted such that *all* satisfy a minimal level. The surprise function approach effectively sums each of the $\alpha - levels$ for each of the constraints, then maximizes this sum with respect to α. Since the optimization is over sets of crisp values coming from fuzzy sets, the surprise approach is a fuzzy optimization method. The salient feature is that surprise uses a dynamic penalty for falling outside distribution/membership values of one. The advantage is that the individual penalties are convex functions, which become infinite as the values approach the endpoints of the support. Moreover, this approach is computationally tractable.

Again, the surprise approach may be used to handle soft constraints of Tanaka, Okuda, and Asai (see [49]) and Zimmermann (see [56]), since these soft constraints can be considered to be fuzzy numbers. However, if soft constraints are handled using surprise functions, the sum of the failure to meet the constraints is minimized rather than forcing each constraint to meet a minimal (fuzzy) feasibility level.

Another historically significant interpretation of fuzzy optimization comes from Verdegay [51], who proposes a method for obtaining a fuzzy solution to a fuzzy problem. This is a deviation from the solutions examined so far, which have been real valued, crisp solutions. Verdegay considers a problem with fuzzy constraints,

$$z = f(x) \tag{22}$$
$$x \subseteq \tilde{C},$$

where the set of constraints have a membership function μ_C, with $\alpha - cut$ $\tilde{C}_\alpha$.

Verdegay defines x_α as the set of solutions that satisfy constraints $\tilde{C}_\alpha$. Then a fuzzy solution to the fuzzy mathematical programming problem is

$$\max_{x \in \tilde{C}_\alpha} z = f(x) \tag{23}$$
$$\forall\, \alpha \in [0,1].$$

Verdegay proposes solving (23) parametrically for $\alpha \in [0,1]$ to obtain a fuzzy solution $\tilde{\chi}$, with $\alpha - cut$ χ_α, which yields fuzzy objective value $\tilde{z}$, with α-cut z_α.

2.4.3 Possibilistic Decision Making Using Possibilistic Optimization

One approach to possibility distributions of parameters ([27] and [32]) allows all constraint violations at an established cost or penalty and minimizes the

expected average, a generalization of expected value ([10], [26], [27], [53], and [54]). This approach considers all possible outcomes as a weighted expected average penalty. The expected average is a type of utility. Another utility is minimizing regret, from Kasperski's article in this volume. This particular utility takes violations as penalties on all outcomes of the constraints. It optimizes over sets of possibility distributions, so it is possibilistic optimization.

Another approach that optimizes over possibility distributions [23] and [24]also optimize over distributions considers constraint feasibility as possibilistic generalizations of *chance constraint* methods. The approach used in [27] and [32] is a possibilistic generalization of the *recourse models* in stochastic optimization (see for example [2]), where violations of constraints are considered as allowable up to a maximum but at a cost. The recourse model in the context of non-probabilistic uncertainty has been studied by [25] where interval parameters/coefficients are treated.

Possibilistic optimization historically was introduced by Buckley [3] as an off-shoot of fuzzy optimization. Because this is a possibilistic linear program, the objective function is governed by a possibilistic distribution. To derive the possibilistic objective function value for a particular solution x, Buckley first specifies the possibility that x satisfies each constraint, and takes the minimum to indicate the possibility that x satisfies all constraints. Buckley next constructs $\text{Poss}[Z = z|x]$, which is the conditional possibility that the objective function Z obtains a particular value z, given values for x. This definition of the possibility distribution motivates Buckley's solution method. Recall that because we are dealing with a possibilistic problem, the solution is governed by a possibilistic distribution. Buckley's method depends upon a static α, chosen *a priori*. The decision maker defines an acceptable level of uncertainty in the objective outcome, $0 < \alpha \leq 1$. For a given α, we define the left and right end-points of the α-cut of a fuzzy number $\tilde{x}$ as $x^-(\alpha)$ and $x^+(\alpha)$, respectively. Using these, Buckley defines a new objective function:

$$Z(\alpha) = c^-(\alpha)x \qquad\qquad (24)$$
$$A^+(\alpha)x \geq b^-(\alpha).$$

This linear program is constrained upon the best-case scenario. That is, for a given α-level, each variable is multiplied by the largest possible coefficient $a_{ij}^+(\alpha)$, and is required to be greater than the smallest possible right-hand side $b_i^-(\alpha)$. We should interpret $Z(\alpha)$ accordingly. If the solution to the linear program is implemented, the possibility that the objective function will attain the level $Z(\alpha)$ is given by α. Stated another way, the best-case scenario is that the objective function attains a value of $Z(\alpha)$, and the possibility of the best case scenario occurring is α.

In the mid 1980s, Tanaka and Asai [47] and Tanaka, Ichahashi, and Asai [50] proposed a technique for dealing with ambiguous coefficients and right-hand sides based upon a possibilistic definition of "greater than zero." First, the objective function is viewed as a goal. As in flexible programming, the goal

becomes a constraint, with the aspiration level for the objective function on the right-hand-side of the inequality. The right-hand sides of the constraints are subtracted so that all the numeric information is contained in a single matrix, which is feasible when it is greater than zero. A fuzzy measure of non-negativity is introduced to gauge optimality of a potential solution x.

Luhanjula's [39] formulation of the possibilistic mathematical program depends upon his concept of "more possible" values. He first defines a possibility distribution Π_X with respect to constraint F as

$$\Pi_X = \mu_F(u),$$

where $\mu_F(u)$ is the degree to which the constraint F is satisfied when u is the value assigned to the solution X. Then the set of more possible values for X, denoted by $V_p(X)$, is given by

$$V_p(X) = \Pi_X^{-1}(\max_u \Pi_X(u)).$$

In other words, $V_p(X)$ contains elements of U which are most compatible with the restrictions defined by Π_X. It follows from intuition and from Luhanjula's formal proof [39] that when Π_X is convex, $V_p(X)$ is a real valued interval, and when Π_X is strongly convex, $V_p(X)$ is a single real number. This formulation varies significantly from the other approaches considered thus far. The possibility of each possibilistic component is maximized individually. Other formulations have required that each possibilistic component $\tilde{c}_j$, $\tilde{A}_{ij}$, and $\tilde{b}_i$ achieve the same possibility level defined by α. This formulation also has a distinct disadvantage over the others presented here. The authors know of no proposed computational method for determining the "more possible" values, V_p, so there appears to be no way to solve the deterministic the problem.

2.4.4   Mixed Fuzzy and Possibilistic Decision Making Using Mixed Possibilistic and Fuzzy Optimization Methods

An optimization problem containing both fuzzy and possibilistic variables is called a *mixed problem* in this chapter. Problems in which one or more possibilistic parameters occurs with one or more fuzzy parameters (or fuzzy inequalities) have been studied (see [24], [36]). Within a quantitative setting, there are two cases for the mixed problem. The first case is a problem that contains both fuzzy and possibilistic parameters (or soft inequalities), but in which each constraint contains exclusively fuzzy or possibilistic parameters. In this case, the fuzzy constraints can be optimized by $\alpha - levels$ (according to [40] or [56]) and the possibilistic constraints by penalized expected averages (according to [27]). For the more complex case in which both fuzzy and possibilistic parameters appear in the same constraint, one approach is to compute the possibilistic distributions of the aggregation of the fuzzy membership functions and optimize over the penalized expected average. The

fuzzy parameter(s) mixed with possibilistic parameter(s) generate a possibilistic distribution that depends on the aggregation operator of the fuzzy membership functions. Another approach is described in Untiedt's chapter in this volume.

3 A Taxonomy of Fuzzy and Possibilistic Optimization for Our Generic Problem

The structure of the generic optimization problem (16) when it is a linear programming problem may be considered to be formed by (i) the *rim* $f(x, c) = c^T x$, and b, e, (ii) the *body* $f(x, a), h(x, d) = Ax$, and (iii) the *relationship*, $\leq, =$ ([21]). For the generic form of the mathematical programming problem (16), we consider a taxonomy based on (i) rim objective function parameters, c, (ii) rim right-hand side parameters, b and e, (iii) body parameters a and d, and (iv) relationship $\leq, =$. For this exposition, a fuzzy/possibilistic optimization problem is considered to be (16) in the presence of data $\{a, b, c, d, e, \leq, =\}$ that is either all or a mixture of real, interval, probabilistic, fuzzy, possibilistic with at least one parameter being fuzzy or possibilistic where soft constraints are assumed to have been translated into fuzzy intervals. If we have a probabilistic optimization problem whose values are known over each $x \in \mathbb{R}$, we would consider it under possibilistic optimization where the upper bound and lower bound (possibility/plausibility and necessity/belief) would be equal. If the probability were known only over sets, then we would have an upper possibility bound and a lower necessity bound as in Figure 3 and do our bound interval-valued possibilistic optimization which is transformed into utility optimization. A right-hand side value that is fuzzy may be interpreted in two ways depending on the context of the problem. First, a fuzzy right-hand side may indicate flexibility. Second, it may indicate (true) decomposable transition modelled by a fuzzy interval. For the former, the constraint becomes a flexible constraint. For the latter, it becomes a possibility.

Note that in the context of interval, fuzzy, possibility, $r \leq s$ and $s \leq r$ does **not** imply $r = s$ as can be seen in the following.

Example 7. *Let*

$$[2, 3]x \leq [3, 6] \ and \tag{25}$$
$$[2, 3]x \geq [3, 6]. \tag{26}$$

The solution of (25) is

$$x = (-\infty, 1].$$

The solution of (26) is

$$x = [3, \infty).$$

Thus (25) and (26) imply that $x = \varnothing$. However,

$$[2,3][\underline{x}, \overline{x}] = [3,6]$$

means that $[\underline{x}, \overline{x}] = [\frac{3}{2}, 2]$.

This example shows that for the constraint fuzzy/possibilistic linear system, $Ax \leq b$ and $Ax \geq b$ are not equivalent to $Ax = b$.

The types of optimization in the presence of interval, fuzzy interval, possibilistic coefficients, and soft constraints are as follows:

1. **Flexible Programming**

   a. Soft constraints relationships $\leq$ and/or $=$ that take on a flexible meaning (come as close as possible or come as close as possible without violation some hard constraints).
   b. The objective function expresses a target desired (come as close as possible to staying under a budgetary value, or attain at least as much profit as was obtained last year).
   c. The right-hand-side value of a constraint is a fuzzy interval which is semantically a target (deliver as close to zero radiation as possible to healthy cells but absolutely to not exceed a critical threshold value).

2. **Utility Programming**

   a. Interval, fuzzy interval, possibilistic cost coefficients of the objective function rim parameter c with real valued coefficient constraint coefficient $a, b, d, e \in \mathbb{R}$.
   b. The objective function rim parameter $c \in \mathbb{R}$ with interval, probability, possibility, fuzzy interval, and one or two of the following - body parameters interval, fuzzy interval, possibilistic a, d and/or rim right-hand values b, e are possibilistic.
   c. Interval-Valued Probability Measure (IVPM) Programming - any of the coefficients a, b, c, d, e may be interval, fuzzy, possibilistic where there may be a mixture of types within one constraint statement.

3. **Random Set** - any of the coefficients a, b, c, d, e may be interval, fuzzy, possibilistic where there may be a mixture of types within one constraint statement.

One also might classify fuzzy and possibilistic programming according to whether or not the solution is a real valued fuzzy interval vector or a real valued vector. Possibilistic programming methods of Buckley [3] and his colleagues and Delgado [5] and his colleagues have considered fuzzy interval-valued solutions. The methods to obtain fuzzy interval solutions are different than those that obtain a real valued solution. Nevertheless, they fall under possibilistic programming or random set programming of the above taxonomy.

References

1. Bellman, R.E., Zadeh, L.A.: Decision-making in a fuzzy environment. Management Science, Serial B 17, 141–164 (1970)
2. Birge, J.R., Louveux, F.: Introduction to Stochastic Programming. Springer, New York (1997)
3. Buckley, J.J.: Solving possibilistic linear programming problems. Fuzzy Sets and Systems 31, 329–341 (1989)
4. de Cooman, G.: Possibility theory I: The measure- and integral - theoretic groundwork. International Journal of General Systems 25(4), 291–323 (1997)
5. Delgado, M., Verdegay, J.L., Vila, M.A.: A general model for fuzzy linear programming. Fuzzy Sets and Systems 29, 21–29 (1989)
6. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics 38, 325–339 (1967)
7. Dubois, D.: Linear programming with fuzzy data. In: Bezdek, J. (ed.) Analysis of Fuzzy Information. Applications in Engineering and Science, vol. III, pp. 241–263. CRC Press, Boca Raton (1897)
8. Dubois, D.: Personal communications (2008, 2009)
9. Dubois, D., Prade, H.: Fuzzy Sets and Systems: Theory and Applications. Academic Press, New York (1980)
10. Dubois, D., Prade, H.: The mean value of a fuzzy number. Fuzzy Sets and Systems 24, 279–300 (1987)
11. Dubois, D., Prade, H.: Possibility Theory. Plenum Press, New York (1988)
12. Dubois, D., Prade, H.: The three semantics of fuzzy sets. Fuzzy Sets and Systems 90, 141–150 (1997)
13. Dubois, D., Prade, H. (eds.): Fundamentals of Fuzzy Sets. Kluwer Academic Press, Dordrecht (2000)
14. Dubois, D., Prade, H.: Formal representations of uncertainty. In: IFSA 2009, Lisboa, Portugal (July 2009)
15. Dubois, D., Karre, E., Mesiar, R., Prade, H.: Fuzzy interval analysis. In: Dubois, D., Prade, H. (eds.) Fundamentals of Fuzzy Sets, ch. 10, pp. 483–581. Kluwer Academic Press, Dordrecht (2000)
16. Dubois, D., Nguyen, H., Prade, H.: Possibility theory, probability theory, and fuzzy sets: Misunderstanding, bridges and gaps. In: Dubois, D., Prade, H. (eds.) Fundamentals of Fuzzy Sets, ch. 7, pp. 343–438. Kluwer Academic Press, Dordrecht (2000)
17. Dubois, D., Prade, H., Sabbadin, R.: Decision-theoretic foundations of qualitative possibility theory. European Journal of Operational Research 128, 459–478 (2001)
18. Fandel, G.: PPS-Systeme: Grundlagen, Methoden, Software, Markanalyse. Springer, Heidelberg (1994)
19. Fishburn, P.C.: Stochastic dominance and moments of distributions. Mathematics of Operations Research 5, 94–100 (1980)
20. Gay, D.M.: Solving interval linear equations. SIAM Journal of Numerical Analysis 19(4), 858–870 (1982)
21. Greenberg, H.: Personal communication (2005)
22. Inuiguchi, M.: Stochastic programming problems versus fuzzy mathematical programming problems. Japanese Journal of Fuzzy Theory and Systems 4(1), 97–109 (1992)

23. Inuiguchi, M., Ichihashi, H., Kume, Y.: Relationships between modality constrained programming problems and various fuzzy mathematical programming problems. Fuzzy Sets and Systems 49, 243–259 (1992)
24. Inuiguchi, M., Sakawa, M., Kume, Y.: The usefulness of possibilistic programming in production planning problems. Int. J. of Production Economics 33, 49–52 (1994)
25. Inuiguchi, M., Tanino, T.: Two-stage linear recourse problems under non-probabilistic uncertainty. In: Yoshida, Y. (ed.) Dynamical Aspect in Fuzzy Decision Making, pp. 117–140. Physica-Verlag, Heidelberg (2001)
26. Jamison, K.D.: Modeling uncertainty using probability based possibility theory with applications to optimization. Ph.D. Thesis, UCD Department of Mathematics (1998)
27. Jamison, K.D., Lodwick, W.A.: Fuzzy linear programming using penalty method. Fuzzy Sets and Systems 119, 97–110 (2001)
28. Jamison, K.D., Lodwick, W.A.: The construction of consistent possibility and necessity measures. Fuzzy Sets and Systems 132, 1–10 (2002)
29. Kaymak, U., Sousa, J.M.: Weighting of constraints in fuzzy optimization. Constraints 8, 61–78 (2003)
30. Klir, G., Yuan, B.: Fuzzy Sets and Fuzzy Logic: Theory and Applications. Prentice Hall, New Jersey (1995)
31. Lodwick, W., Bachman, K.: Solving Large Scale Fuzzy Possibilistic Optimization Problems. Fuzzy Optimization and Decision Making 4(4), 257–278 (2005)
32. Lodwick, W.A., Jamison, K.D.: A computational method for fuzzy optimization. In: Ayyub, B., Gupta, M. (eds.) Uncertainty Analysis in Engineering and Sciences: Fuzzy Logic, Statistics, and Neural Network Approach, ch. 19. Kluwer Academic Publishers, Dordrecht (1997)
33. Lodwick, W.A., Jamison, K.D.: Estimating and validating the cumulative distribution of a function of random variables: Toward the development of distribution arithmetic. Reliable Computing 9, 127–141 (2003)
34. Lodwick, W.A., Jamison, K.D.: Theory and semantics for fuzzy and possibilistic optimization. In: Fuzzy Logic, Soft Computing and Computational Intelligence (Eleventh International Fuzzy Systems Association World Congress), Beijing, China, July 28-31, vol. III, pp. 1805–1810 (2005)
35. Lodwick, W.A., Jamison, K.D.: Theory and semantics for fuzzy and possibilistic optimization. Fuzzy Sets and Systems 158(7), 1861–1871 (2007)
36. Lodwick, W.A., Jamison, K.D.: The Use of interval-valued probability measure in optimization under uncertainty for problems containing a mixture of possibilistic, probabilistic and interval uncertainty. In: Melin, P., Castillo, O., Aguilar, L.T., Kacprzyk, J., Pedrycz, W. (eds.) IFSA 2007. LNCS (LNAI), vol. 4529, pp. 361–370. Springer, Heidelberg (2007)
37. Lodwick, W.A., Jamison, K.D.: Interval-Valued Probability in the Analysis of Problems Containing a Mixture of Possibilistic, Probabilistic, and Interval Uncertainty. Fuzzy Sets and Systems 159(1), 2845–2858 (2008)
38. Lodwick, W.A., Neumaier, A., Newman, F.D.: Optimization under uncertainty: methods and applications in radiation therapy. In: Proceedings 10th IEEE International Conference on Fuzzy Systems, vol. 3, pp. 1219–1222 (2001)
39. Luhandjula, M.K.: On possibilistic linear programming. Fuzzy Sets and Systems 18, 15–30 (1986)
40. Neumaier, A.: Fuzzy modeling in terms of surprise. Fuzzy Sets and Systems 135(1), 21–38 (2003)

41. Puri, M.L., Ralescu, D.: Fuzzy measures are not possibility measures. Fuzzy Sets and Systems 7, 311–313 (1982)
42. Rommelfanger, H.J.: The advantages of fuzzy optimization models in practical use. Fuzzy Optimization and Decision Making 3, 295–309 (2004)
43. Rommelfanger, H.J.: Personal communication (April 2009)
44. Russell, B.: Vagueness. Australasian Journal of Psychology and Philosophy 1, 84–92 (1923)
45. Shafer, G.: Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
46. Simon, H.: The Sciences of the Artificial, 2nd edn. The MIT Press, Cambridge (1969/1981)
47. Tanaka, H., Asai, K.: Fuzzy linear programming with fuzzy numbers. Fuzzy Sets and Systems 13, 1–10 (1984)
48. Tanaka, H., Okuda, T., Asai, K.: On fuzzy mathematical programming. Transactions of the Society of Instrument and Control Engineers 9(5), 607–613 (1973) (in Japanese)
49. Tanaka, H., Okuda, T., Asai, K.: On fuzzy mathematical programming. J. of Cybernetics 3, 37–46 (1974)
50. Tanaka, H., Ichihashi, H., Asai, K.: Fuzzy Decision in linear programming with trapezoid fuzzy parameters. In: Kacpryzk, J., Yager, R.R. (eds.) Management Decision Support Systems Using Fuzzy Sets and Possibility Theory, pp. 146–154. Verlag TUV, Koln (1985)
51. Verdegay, J.L.: Fuzzy mathematical programming. In: Gupta, M.M., Sanchez, E. (eds.) Fuzzy Information and Decision Processes, pp. 231–237. North Holland Company, Amsterdam (1982)
52. Von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press, Princeton (1947)
53. Yager, R.: On choosing between fuzzy subsets. Kybernetes 9, 151–154 (1980)
54. Yager, R.: A procedure for ordering fuzzy subsets of the unit interval. Information Sciences 24, 143–161 (1981)
55. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1, 3–28 (1978)
56. Zimmermann, H.: Description and optimization of fuzzy systems. International J. of General Systems 2, 209–215 (1976)