FOURIER SERIES APPROXIMATIONS OF LIKELIHOOD-BASED FUZZY SETS

Isaac Rudnick

Computer Science Department, California Polytechnic State University, San Luis Obispo, United States

ABSTRACT

By Zadeh's original formulation, likelihood distributions can be seen as a unique kind of fuzzy set whose logical operations have meaningful probabilistic interpretations. In this work, we develop a variant of this fuzzy set in which an arbitrary number of fuzzy logic operations may be applied without increasing the space and time required for membership evaluation. By using a band-limited Fourier series approximation for a truncated Gaussian kernel, we also demonstrate a significant reduction in time and space requirements for mixture model evaluation. Probabilistic and fuzzy set interpretations, as well as benchmarks against Standard Additive Models (SAMs) are provided along with an analysis of complexity and scaling properties

KEYWORDS

Fuzzy Sets, Fuzzy Logic, Standard Additive Models, Gaussian Mixture Models, Fourier Series, Fourier Basis Density Model.

1. INTRODUCTION

Fuzzy sets and fuzzy logic operations have long represented helpful means of expressing intuitive set membership information for different concepts. At the same time, probabilistic models have provided a mathematically rigorous framework for representing uncertainty, learning from data, and updating beliefs through observed evidence.

In this paper, we show that continuous approximations of likelihood distributions and their conjugate priors represent a unique family of fuzzy sets with meaningful interpretations. By viewing likelihood functions as continuous membership relations over a compact domain, standard fuzzy logical operations can be interpreted directly in probabilistic terms.

Under this view, negation corresponds to complementing belief mass, intersection corresponds to joint likelihood under independence, and union corresponds to the likelihood of at least one event occurring. This gives fuzzy set operations a more concrete statistical semantics.

For computational efficiency, we develop a Fourier series-based representation of these continuous likelihood approximations. By approximating truncated kernel functions through truncated Fourier series, fuzzy operations can be redefined on harmonic coefficients. This yields a compact and analytically tractable representation in which composition does not increase the dimensionality of the underlying approximation space. Repeated applications of fuzzy logical operators in this space are thus more scalable than Standard Additive Models (SAMs) of fuzzy logic.

David C. Wyld et al. (Eds): WiMNeT, NLDM, EDUPT, SIP, AISO – 2026
pp. 83-98, 2026. CS & IT - CSCP 2026 DOI: 10.5121/csit.2026.1601008

To illustrate our approach, we study truncated Gaussian kernels as approximations of skewed distributions and derive analytic expressions for their Fourier coefficients in terms of the error function. We then use these coefficients to define fuzzy set operations over finite harmonic expansions and examine their behavior in a kernel aggregation setting.

Particular attention is given to the issue of vanishing magnitude under repeated intersections, for which we introduce a renormalization procedure that preserves signal strength without abandoning the fuzzy-logical interpretation.

Together, these Fourier series approximations of likelihood-based fuzzy sets may offer a useful bridge between interpretable fuzzy reasoning and scalable probabilistic modeling.

2. Related Work

2.1. Fuzzy Sets

Fuzzy sets and fuzzy logic have historically offered a straightforward means of expressing intuitive set membership information for various concepts. The concept of a "fuzzy set" was originally formalized by Zadeh, but it quickly expanded into "fuzzy logic" [1]. Fuzzy logic systems drew a significant amount of attention to the area due to their satisfaction of the universal approximation theorem while retaining interpretable intermediate representations. The usefulness of these properties was demonstrated with the advent of Mamdani-style controllers [2] [3]. Unfortunately, the combinatory computational complexity of evaluating traditional fuzzification maps and their composition under fuzzy logical operations limited their scalability [3] [4]. This limited computational efficiency and expressiveness led to their being superseded by more powerful digital control systems, graph representations, and probabilistic models [3]. Nevertheless, the framework has seen renewed interest for knowledge graph embedding [5] and lexical relation models [6]. Although these ontology-based embedding composition schemes have vastly reduced the manual cost of designing fuzzy systems' composition rules, the computational complexity remains a bottleneck for their broader adoption.

2.2. Fuzzy Priors

Previous attempts at reconciling probability-theoretic models of uncertainty with soft and fuzzy logic notions of uncertainty have been quite varied but have generally been met with limited practical success. In his 2008 paper, Glen Meeden proposed decision-theoretic, regression-based fuzzy set construction from prior distributions. He proposes a generalizable map from a prior distribution to their most representational fuzzy set. Unfortunately, this mapping is not injective, with many possible priors for a given fuzzy set [7]. Another caveat of this loss-based approach to fuzzy set generation is the ambiguity of fuzzy set operations' interpretations with respect to possible prior distributions.

Later work by Osoba, Mitaim, and Kosko offers a much more rigorous, analytic framework for fuzzy sets' construction and relevance to approximating prior, hyperprior, and non-conjugate posterior distributions. In their work, they also provide loss functions for tuning fuzzy Standard Additive Model (SAM) parameters to approximate these distributions [8]. Similar to the work Meeden, the authors also omit interpretations of fuzzy set operations' effects on prior and posterior distributions [7] [8].

SAMs also suffer from exponential rule explosion in high dimensions and tuning such fuzzy rules compounds this computational complexity, making systems with "even four independent variables intractable in practice" [4]. The space requirements and evaluation times for SAMs also scale linearly with the number of distributions involved. For low-dimensional distributions with few modes, this can be done with reasonable efficiency. But as sets' dimensionalities and quantity of peaks increases, such models tend to break down [8] [4].

2.3. Continuous Approximations of Skewed Likelihood Distributions

Discrete likelihood distributions can generally be approximated by aggregating a collection of binomial distributions. This preserves certain features that make them a useful category of fuzzy set and is why continuous approximations are useful. The conjugate prior and posterior distributions of the binomial distribution (both Beta distributions) can be approximated using a similar kernel. Any space of continuous distributions that faithfully approximates these three distributions, however, must support some amount of skewedness.

There are various existing continuous approximations of the Binomial distribution that double as prior and posterior approximators, so we consider the 3 most popular: Poisson, normal, and beta distributions.

Poisson approximations are reasonable Binomial distribution approximators when the prior $p \sim 0$ and the sample size n is very large, but they fall short for large priors $p \sim 1$. While this can sometimes be mitigated by approximating the likelihood distribution of $1 - p$ instead, this alternative can complicate its use as a kernel in mixture models.

Normal approximations are more versatile and indeed more common in mixture models. However, they lack skew information, making them effective primarily for mid-range priors $p \sim 0.5$. Although this can be mitigated with bounding (truncating) and rescaling, the result is still a worse approximation than a Beta distribution.

Beta distributions represent both the conjugate prior and conjugate posterior of the Binomial likelihood distribution [8]. The Beta distribution is also a good approximator of Binomial distributions with arbitrary priors because of its flexible shape, which can represent both double-peaked and single-peaked distributions, unlike a Gaussian [9]. Beta distributions' aggregation in mixture models also produce similar results as Gaussian kernels when $p \sim 0.5$ and support more faithful aggregation of heavily skewed distributions where $p(1 - p) \sim 0$. This makes them the most mathematically sound kernel for prior, likelihood, and posterior distributions. Unfortunately, evaluating the Beta distribution becomes computationally intractable for arbitrary levels of precision, significantly reducing their usefulness for kernel generation in mixture models [10]. Thankfully, a truncated and rescaled Gaussian serves as a reasonable approximator. We use this fact to construct a Fourier series kernel approximation in our analysis.

2.4. Fourier Basis Density Models

In their 2024 paper, De la Fuente, Singh, and Balle introduce a lightweight probability density model. This Fourier Basis Density Model (FBM) is parameterized by a constrained Fourier basis. They evaluated this model on a toy compression task and found that it outperformed a deep factorized model with similar computational overhead. Additional analysis showed significantly lower KL-divergence compared with gaussian mixture and deep factorization models with the same pa- rameter counts [11]. Our work draws on many of FBMs' design choices with a focus on fuzzy logic operations' effects on fuzzy Fourier likelihood sets. As will be discussed in section 5.4.1, series' coefficients may be rescaled by the 0th harmonic term to get a kernel density estimate for a Gaussian Mixture Model (GMM) [11].

3. Fuzzy Set Operations On Likelihood Distributions

Consider two independent binary random variables p and q. Suppose the probability of a positive sample from p is P(p) and q is defined similarly. Since these variables form Bernoulli distributions, their negative outcomes will occur with probabilities 1−P(p) and 1−P(q) respectively. Noting that the joint probability of two independent random variables is given by the product of their probabilities, we describe the joint probability distribution of states of p and q as

p	q	$\neg p$	$\neg q$	$P(\text{state})$
F	F	T	T	$P(\neg p)P(\neg q)$
F	T	T	F	$P(\neg p)P(q)$
T	F	F	T	$P(p)P(\neg q)$
T	T	F	F	$P(p)P(q)$

So we have that:

$$P(\neg p)=1-P(p) \quad (1)$$

$$P(p\wedge q)=P(p)P(q) \quad (2)$$

$$P(p\vee q)=P(\neg p)P(q)+P(p)P(\neg q)+P(p)P(q) \quad (3)$$

$$= (1-P(p))P(q)+P(p)(1-P(q))+P(p)P(q) \quad (4)$$

$$=P(q)-P(p)P(q)+P(p)-P(p)P(q)+P(p)P(q) \quad (5)$$

$$=P(q)-P(p)P(q)+P(p) \quad (6)$$

$$=P(p)+P(q)-P(p)P(q) \quad (7)$$

If we treat each Yn in a probability mass function PMF(Yn) as such a binary random variable, then the joint probability of two such likelihood distributions is defined by their product. Extrapolating to our continuous approximation of the likelihood distribution, then, requires analogous operations.

Negation: 1− f(x) as defined by Zadeh [1] (8)
Intersection: f(x)g(x) as defined by Dubois and Prade [12] (9)
Union: f(x) +g(x)− f(x)g(x) as defined by Dubois and Prade [12] (10)

As defined above, the continuous PMF(Yn) represents a fuzzy set, therefore let f(x) and g(x) denote similar continuous likelihood distribution approximations on a sampling rate given by x. Since f(x) and g(x) meet the criteria to be considered fuzzy sets [1], we unify previous works' fuzzy set operator definitions with the likelihood distribution interpretation above:

Within this framework, then, the fuzzy negation describes the likelihood of an event not occurring, fuzzy intersection describes the joint likelihood distribution, and fuzzy union describes the likelihood of at least one event occurring.

4. FUZZY FOURIER LIKELIHOOD SETS

Consider 2 fuzzy sets $f(x)$ and $g(x)$ with domains $[0,1] \subset \mathbb{R}$. If we treat these as continuous time signals $f(t)$ and $g(t)$ for $t \in [0,1]$, then their Fourier transforms under different fuzzy set operations are then described below:

$$\widehat{f}(k) = \int_0^1 f(t)e^{-2\pi ikt}\,dt \tag{11}$$

$$\widehat{\neg f}(k) = \begin{cases} -\widehat{f}(k) & \text{if } k \neq 0 \\ 1-\widehat{f}(k) & \text{if } k = 0 \end{cases} \tag{12}$$

$$\widehat{f \cap g}(k) = (\widehat{f} * \widehat{g})(k) \tag{13}$$

$$\widehat{f \cup g}(k) = \widehat{f}(k) + \widehat{g}(k) - (\widehat{f} * \widehat{g})(k) \tag{14}$$

In many cases, approximating continuous time signals' Fourier transforms is done most efficiently using the Fast Fourier Transform (FFT). In kernel-based models, however, the time signal associated with the kernel function is known ahead of time. This makes an analytic approach to calculating kernels' harmonic coefficients more appealing. By approximating the coefficients with an explicit formula, we are able to take advantage of existing methods for evaluating a range of their harmonics (particularly error function calculations [13]).

We therefore evaluate the L-periodic Fourier series approximations of $f(t)$ and $g(t)$ using their transformed harmonics in the range $[-n,n]$. Doing so is then as simple as using their discrete harmonics as coefficients in the series expansion. In essence,

$$f(t) \sim \frac{1}{L}\sum_{k=-n}^{n} e^{\frac{2\pi ikt}{L}} \widehat{f}\left(\frac{k}{L}\right) \tag{15}$$

$$\neg f(t) \sim \frac{1}{L}\sum_{k=-n}^{n} e^{\frac{2\pi ikt}{L}} \widehat{\neg f}\left(\frac{k}{L}\right) \tag{16}$$

$$f(t) \cap g(t) \sim \frac{1}{L}\sum_{k=-n}^{n} e^{\frac{2\pi ikt}{L}} \widehat{f \cap g}\left(\frac{k}{L}\right) \tag{17}$$

$$f(t) \cup g(t) \sim \frac{1}{L}\sum_{k=-n}^{n} e^{\frac{2\pi ikt}{L}} \widehat{f \cup g}\left(\frac{k}{L}\right) \tag{18}$$

$$\tag{19}$$

5. FUZZY SET OPERATIONS ON FUZZY FOURIER LIKELIHOOD SETS

By approximating a truncated Gaussian with a Fourier series, we can represent fuzzy logic operations as transformations on the Hilbert space in which the approximations' harmonics lie. To better understand this, consider the Fourier series approximations for fuzzy sets $f(x)$ and $g(x)$ whose coefficients are given by $\widehat{f}(k)$ and $\widehat{g}(k)$ respectively.

Let $K = [-n, 1-n, 2-n, \ldots, -1, 0, 1, \ldots, n-2, n-1, n]$ be an index of harmonics in $[-n,n]$.

Let $\widehat{F}, \widehat{G} \in \mathbb{C}^{2n+1}$ denote vectors in $\mathbb{C}^2 n$ whose components map to the coefficients of $f(x)$ and $g(x)$ respectively, as indexed by K.

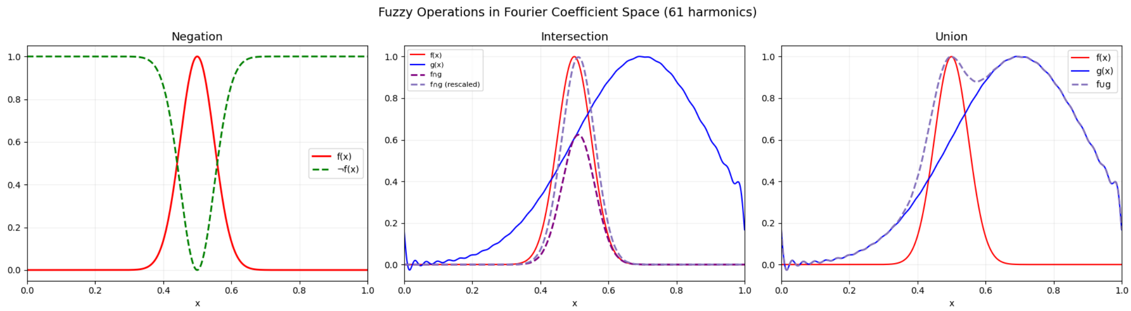

Figure 1. Examples of fuzzy set negation [left], intersection with/without rescaling [center], and union [right] applied to fuzzy Fourier sets as defined in section 6. The mean and variance of each kernel are given in figure 3

5.1. Negation

Since equation 12 represents an element-wise operation on the coefficients of $\widehat{f}(k)$, $\widehat{\neg F}$ is defined by

$$\widehat{\neg F} = [F_{-n}, \ldots, F_{-1}, 1 - F_0, F_1 \ldots, -F_n] \tag{20}$$

The result of this can be seen in 1

5.2. Intersection

Equation 13 defines the fuzzy set intersection of $f(t)$ and $g(t)$ as their convolution in the frequency domain. Since F and G correspond to discrete harmonics of $\widehat{f}(t)$ and $\widehat{g}(t)$, we define their element-wise convolution as:

$$\widehat{f \cap g}(k) = (\widehat{f} * \widehat{g})(k) = \sum_{l=-n}^{n} \widehat{f}(l)\widehat{g}(k-l)$$

This lets us define the convolution in terms of F and G as

$$\widehat{F \cap G} = \widehat{F} * \widehat{G} = \left[\sum_{l=-n}^{n} \widehat{f}(l)\widehat{g}(-n-l), \ldots, \sum_{l=-n}^{n} \widehat{f}(l)\widehat{g}(n-l) \right]$$

For 2 series expansions with $2n+1$ coefficients each, a naive approach to convolution would evaluate the coefficients of the intersection in $O((2n+1)^2)$ time.

Multiplying the signals in the time domain and mapping their result back to the frequency domain is reasonable, but using the Fast Fourier Transform can lose information between the spaces.

An efficient compromise might involve constructing a Toeplitz matrix for the various offsets of $\widehat{g}$, call it $G' \in M_{(2n+1)\times(2n+1)}(\mathbb{C})$, and multiplying it with F. An element-wise view of this operation can be seen in Figure 2

Constructing G' as defined above would take roughly $O((2n+1)^2)$ time and $O((2n+1)^2)$ space, but the matrix multiplication it supports with F would take roughly $O((2n+1)^2)$ time.

Due to time constraints, we used the naive approach (implemented with numpy) to generate figures and leave more efficient implementations as future work. An example of this operation can be seen in figure 1

$$G' = \begin{bmatrix} \hat{g}(0) & \hat{g}(-1) & \hat{g}(-2) & \cdots & \hat{g}(-2n) \\ \hat{g}(1) & \hat{g}(0) & \hat{g}(-1) & \cdots & \hat{g}(-2n+1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{g}(2n-1) & \hat{g}(2n-2) & \hat{g}(2n-3) & \cdots & \hat{g}(-2n+2) \\ \hat{g}(2n) & \hat{g}(2n-1) & \hat{g}(2n-2) & \cdots & \hat{g}(0) \end{bmatrix}.$$

$$\begin{bmatrix} \sum_{l=-n}^{n} \hat{f}(l)\,\hat{g}(-n-l) \\ \sum_{l=-n}^{n} \hat{f}(l)\,\hat{g}(-n+1-l) \\ \vdots \\ \sum_{l=-n}^{n} \hat{f}(l)\,\hat{g}(n-1-l) \\ \sum_{l=-n}^{n} \hat{f}(l)\,\hat{g}(n-l) \end{bmatrix} = \begin{bmatrix} \hat{g}(0) & \hat{g}(-1) & \cdots & \hat{g}(-2n) \\ \hat{g}(1) & \hat{g}(0) & \cdots & \hat{g}(-2n+1) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{g}(2n-1) & \hat{g}(2n-2) & \cdots & \hat{g}(-1) \\ \hat{g}(2n) & \hat{g}(2n-1) & \cdots & \hat{g}(0) \end{bmatrix} \begin{bmatrix} \hat{f}(-n) \\ \hat{f}(-n+1) \\ \vdots \\ \hat{f}(n-1) \\ \hat{f}(n) \end{bmatrix}.$$

Figure 2. The Toeplitz matrix for G' and an element view of the terms of the equality $F * G = G'F$.

5.3. Union

Using the definition for element-wise fuzzy set union in the frequency domain given by equation 14, we find

$$\widehat{F \cup G} = \widehat{F} + \widehat{G} - \widehat{F} * \widehat{G} \tag{21}$$

Since our definition of two fuzzy sets' union involves their intersection, the time complexity of evaluation is also $O((2n+1)^2)$. An example implementation is shown in figure 1

5.4. Renormalization and Rescaling

5.4.1. Rescaling

A common issue with fuzzy set intersection is the potential for a vanishing limit. When fuzzy intersection is applied iteratively without appropriate rescaling, the maximum value attained by $(f \cap g)(t)$ may eventually diminish to zero. While consistent with Zadeh's original formulation, this behavior is often undesirable for machine learning (ML) tasks where maintaining signal strength is important.

To address this, we rescale the signal in the frequency domain. To renormalize the intersection of two fuzzy Fourier sets $(f \cap g)(t)$ and prevent it from vanishing, we divide the coefficients of their series approximations by the sum of their absolute values, as described by equation 22.

$$\frac{\widehat{F \cap G}}{\sum_{k=-n}^{n} \left| \left(\widehat{F \cap G} \right)_k \right|} \tag{22}$$

An example of this rescaling can be seen in figure 1

This rescaling property, coupled with the fixed dimensionality of composed kernel density estimates, may have great potential for defining fixed-complexity loss gradients. Such gradients may be defined manually by rescaling fuzzy Fourier likelihood sets with respect to the smallest 0th harmonic and subsequently composing them with fuzzy set operations. The final fuzzy set would just require renormalization to represent a valid probability density function.

5.4.2. Renormalization

In mixture model contexts, it's more useful to renormalize a fuzzy Fourier likelihood set into a kernel density estimate. Thankfully, if we recall equation 11, the 0th harmonic (i.e. the DC term's coefficient) of F is given by

$$\int_0^1 f(t)\, dt$$

This represents the cumulative likelihood mass, so the probability density associated with some set F is found by dividing all coefficients by this term.

6. Case Study: A Skewed Kernel Approximator for Gaussian Mixture Models

To construct a series approximation for a skewed kernel, we started with a truncated Gaussian approximation for a Beta likelihood distribution. An example of 2 kernels constructed with this approximation can be seen in 3.

6.1. Series Approximation Derivation

The Normal distribution is often used to approximate the Binomial distribution in cases where $p \sim 0.5$ and can even be truncated and rescaled to approximate Poisson and Beta distributions.

The mean μ and variance σ^2 for approximating a Beta distribution with parameters α and β are given by:

$$\mu = \frac{\alpha}{\alpha + \beta} \tag{23}$$

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{24}$$

$$\tag{25}$$

and the truncated gaussian is thus given by

$$f(t) = \begin{cases} e^{-\frac{(t-\mu)^2}{2\sigma^2}} & \text{if } t \in [a,b] \\ 0 & \text{if } t \notin [a,b] \end{cases}$$

The Fourier transform of the truncated Gaussian defined on $[a,b]$ is then

$$\widehat{f}(k) = \int_a^b e^{-\frac{(t-\mu)^2}{2\sigma^2}} e^{-2\pi ikt}\,dt$$

This can be further rewritten to express $\widehat{f}(k)$ analytically in terms of the error function on the interval $[a,b]$,

$$\widehat{f}(k) = \sigma\sqrt{\frac{\pi}{2}}e^{2\pi ik(-\mu+\pi i\sigma^2 k)}\left[\operatorname{erf}\left(\frac{b-\mu+2\pi i\sigma^2 k}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{a-\mu+2\pi i\sigma^2 k}{\sqrt{2}\sigma}\right)\right] \tag{26}$$

Now let

$$L = b - a \tag{27}$$

$$c_k = \frac{1}{L}\widehat{f}\left(\frac{k}{L}\right) \tag{28}$$

$$z_k = \frac{-\mu L + 2\pi i\sigma^2 k}{\sqrt{2}\sigma L} \tag{29}$$

For the L-periodic complex Fourier series approximating $f(x)$ on $[a,b]$ is thus

$$f(x) \sim \sum_{k\in\mathbb{Z}} c_k e^{\frac{2\pi inx}{L}}$$

By substituting in our definitions of c_k and z_k, we get

$$c_k = \frac{\sigma}{L}\sqrt{\frac{\pi}{2}}e^{-\frac{2\pi ik\mu}{L}-\frac{2\pi^2\sigma^2k^2}{L^2}}\left[\operatorname{erf}\left(\frac{b}{\sqrt{2}\sigma}+z_k\right) - \operatorname{erf}\left(\frac{a}{\sqrt{2}\sigma}+z_k\right)\right] \tag{30}$$

Hence the n-harmonic Fourier approximation is

$$f_n(x) = \sum_{k=-n}^{n}\frac{\sigma}{L}\sqrt{\frac{\pi}{2}}e^{-\frac{2\pi ik\mu}{L}-\frac{2\pi^2\sigma^2k^2}{L^2}}\left[\operatorname{erf}\left(\frac{b}{\sqrt{2}\sigma}+z_k\right) - \operatorname{erf}\left(\frac{a}{\sqrt{2}\sigma}+z_k\right)\right]e^{\frac{2\pi ikx}{L}} \tag{31}$$

To get an approximation for the truncated gaussian on $[0,1]$ with a period of 1, then, we have the following:

$$c_k = \frac{1}{1}\widehat{f}\left(\frac{k}{1}\right) \tag{32}$$

$$= \widehat{f}(k) \tag{33}$$

$$= \frac{\sigma}{1}\sqrt{\frac{\pi}{2}}e^{-\frac{2\pi ik\mu}{1}-\frac{2\pi^2\sigma^2k^2}{1^2}}\left[\operatorname{erf}\left(\frac{1}{\sqrt{2}\sigma}+z_k\right) - \operatorname{erf}\left(\frac{0}{\sqrt{2}\sigma}+z_k\right)\right] \tag{34}$$

$$= \sigma\sqrt{\frac{\pi}{2}}e^{2\pi ik(-\mu+\pi i\sigma^2 k)}\left[\operatorname{erf}\left(\frac{1}{\sqrt{2}\sigma}+z_k\right) - \operatorname{erf}(z_k)\right] \tag{35}$$

$$= \sigma\sqrt{\frac{\pi}{2}}e^{2\pi ik(-\mu+\pi i\sigma^2 k)}\left[\operatorname{erf}\left(\frac{1-\mu+2\pi i\sigma^2 k}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{-\mu+2\pi i\sigma^2 k}{\sqrt{2}\sigma}\right)\right] \tag{36}$$

An example of this series approximation for normal distributions with $\mu = 0.5, 0.7$ and $\sigma = 0.05, 0.2$ can be seen in Figure 3

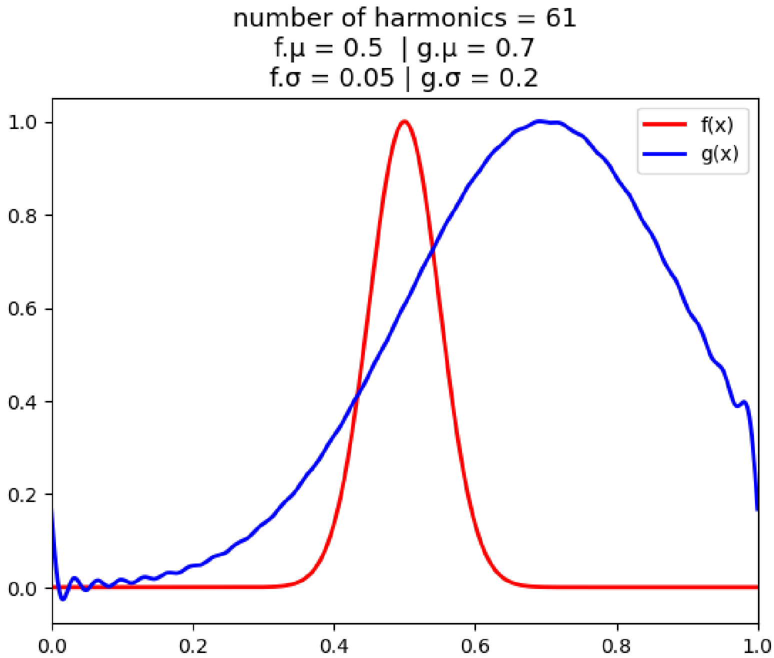

Figure 3. 61-term Fourier series approximations for 2 truncated Gaussians centered at 0.5 and 0.7 with variances of 0.05 and 0.2 respectively.

6.2. Numerical stability

Due to its greater numerical stability, we rewrite equation 35 in terms of the scaled complementary error function with the following substitutions

$$\operatorname{erf}(x) = 1 - e^{-x^2}\operatorname{erfcx}(x) \tag{37}$$

$$a = -z_k^2 \tag{38}$$

$$b = -\left(\frac{1}{\sqrt{2}\sigma} + z_k\right)^2 \tag{39}$$

This gives

$$c_k = \sigma\sqrt{\frac{\pi}{2}}e^{2\pi ik(-\mu+\pi i\sigma^2 k)}\left[\operatorname{erf}\left(\frac{1}{\sqrt{2}\sigma} + z_k\right) - \operatorname{erf}(z_k)\right] \tag{40}$$

$$= \sigma\sqrt{\frac{\pi}{2}}e^{2\pi ik(-\mu+\pi i\sigma^2 k)}\left[\left(1 - e^b\operatorname{erfcx}\left(\frac{1}{\sqrt{2}\sigma} + z_k\right)\right) - (1 - e^a\operatorname{erfcx}(z_k))\right] \tag{41}$$

$$= \sigma\sqrt{\frac{\pi}{2}}e^{2\pi ik(-\mu+\pi i\sigma^2 k)}\left[e^a\operatorname{erfcx}(z_k) - e^b\operatorname{erfcx}\left(\frac{1}{\sqrt{2}\sigma} + z_k\right)\right] \tag{42}$$

To simplify this further, consider expanding a as follows

$$
\begin{aligned}
a &= -z_k^2 && (43)\\
&= -\left(\frac{-\mu L + 2\pi i\sigma^2 k}{\sqrt{2}\,\sigma L}\right)^2 && (44)\\
&= -\frac{\left(-\mu L + 2\pi i\sigma^2 k\right)^2}{2\,\sigma^2 L^2} && (45)\\
&= -\frac{\mu^2 L^2 - 4\pi i\mu\sigma^2 Lk - 4\pi^2\sigma^4 k^2}{2\,\sigma^2 L^2} && (46)\\
&= \frac{4\pi i\sigma^2 k\left(\mu L - \pi i\sigma^2 k\right) - \mu^2 L^2}{2\,\sigma^2 L^2} && (47)\\
&= \frac{2\pi ik\left(\mu L - \pi i\sigma^2 k\right)}{L^2} - \frac{\mu^2}{2\,\sigma^2} && (48)
\end{aligned}
$$

So for $L = 1$,

$$a = 2\pi ik\left(\mu - \pi i\sigma^2 k\right) - \frac{\mu^2}{2\,\sigma^2}$$

Similarly, consider expanding b in equation 42 and note

$$
\begin{aligned}
b &= -\left(\frac{1}{\sqrt{2}\,\sigma} + z_k\right)^2 && (49)\\
&= a - \frac{\sqrt{2}}{\sigma} z_k - \frac{1}{2\sigma^2} && (50)\\
&= 2\pi ik\left(\mu - \pi i\sigma^2 k\right) - \frac{\mu^2}{2\,\sigma^2} - \frac{\sqrt{2}}{\sigma}\left(\frac{-\mu L + 2\pi i\sigma^2 k}{\sqrt{2}\,\sigma L}\right) - \frac{1}{2\sigma^2} && (51)\\
&= 2\pi ik\left(\mu - \pi i\sigma^2 k\right) - \frac{\mu^2}{2\,\sigma^2} - \frac{-\mu + 2\pi i\sigma^2 k}{\sigma^2} - \frac{1}{2\sigma^2} && (52)\\
&= 2\pi ik\left(\mu - \pi i\sigma^2 k\right) - \frac{\mu^2}{2\,\sigma^2} + \frac{\mu}{\sigma^2} - 2\pi ik - \frac{1}{2\sigma^2} && (53)\\
&= 2\pi ik\left(\mu - \pi i\sigma^2 k\right) - 2\pi ik - \frac{(1-\mu)^2}{2\sigma^2} && (54)
\end{aligned}
$$

Distributing $e^{2\pi ik(-\mu+\pi i\sigma^2 k)}$ to the erfcx coefficients is equivalent to adding their exponents, yielding

$$
\begin{aligned}
& 2\pi ik(-\mu + \pi i\sigma^2 k) - z_k^2 && (55)\\
&= 2\pi ik(-\mu + \pi i\sigma^2 k) + 2\pi ik\left(\mu - \pi i\sigma^2 k\right) - \frac{\mu^2}{2\,\sigma^2} && (56)\\
&= 2\pi ik(-\mu + \pi i\sigma^2 k\mu - \pi i\sigma^2 k) - \frac{\mu^2}{2\,\sigma^2} && (57)\\
&= -\frac{\mu^2}{2\,\sigma^2} && (58)
\end{aligned}
$$

Similarly,

$$2\pi ik(-\mu+\pi i\sigma^2 k)-\left(\frac{1}{\sqrt{2}\sigma}+z_k\right)^2 \tag{59}$$

$$=2\pi ik(-\mu+\pi i\sigma^2 k)+2\pi ik\left(\mu-\pi i\sigma^2 k\right)-2\pi ik-\frac{(1-\mu)^2}{2\sigma^2} \tag{60}$$

$$=-2\pi ik-\frac{(1-\mu)^2}{2\sigma^2} \tag{61}$$

Simplifying equation 42 and with our reduced substitutions thus gives

$$c_k=\frac{\sigma}{L}\sqrt{\frac{\pi}{2}}\left[e^{-\frac{\mu^2}{2\sigma^2}}\text{erfcx}\left(\frac{-\mu+2\pi i\sigma^2 k}{\sqrt{2}\sigma}\right)-e^{-2\pi ik-\frac{(1-\mu)^2}{2\sigma^2}}\text{erfcx}\left(\frac{1-\mu+2\pi i\sigma^2 k}{\sqrt{2}\sigma}\right)\right] \tag{62}$$

We therefore use this numerically stable formulation of c_k in our case study implementation.

6.3. Set Operation Complexity Analysis

By approximating a truncated Gaussian with a Fourier series, fuzzy logic operations can be represented as transformations on the Hilbert space where the approximations' harmonics reside. Applying the fuzzy set composition methods described above to a collection of kernels therefore retains constant function space dimensionality with respect to the number of kernels aggregated. Figure 1 highlights the benefits of fuzzy set operations applied to fuzzy Fourier likelihood distributions $P_i, P_j \in \mathbb{H}$ within this framework.

Model	$O(\neg P_i)$	$O(P_i \cap P_j)$	$O(P_i \cup P_j)$	$\dim(P_i \cap P_j)$	$\dim(P_i \cup P_j)$
Standard Additive Models	$c_i c_j$	$c_i c_j$	$c_i + c_j$	$c_i + c_j$	$c_i + c_j$
Fuzzy Fourier Sets	d	d^2	d^2	d	d

Table 1. A comparison between theoretic time complexities of set operations in different fuzzy set composition frameworks. P_i and P_j represent fuzzy sets in P. d denotes the dimensionality of the Hilbert space within which the kernel functions lie. c_i denotes the dimensionality of the Hilbert space in which P_i resides if it may be different from d.

6.4. Mixture Model Evaluation

To assess the usefulness our fuzzy set formulation, we ran several benchmarks with variable spectral bandwidths and kernel counts. For our baseline, we truncated the result of applying

```
scipy.stats.norm.pdf
```

to arrays of randomly generated means and variances in batch form. All evaluation was conducted on a CPU in Google Colab with no GPU acceleration. The mean of 200 mixture evaluation times was used for each kernel count.

As seen in figure 4, evaluation time scales more or less linearly with harmonic count. Fuzzy Fourier Set (FFS)-based mixture models' log-likelihood and Kullback-Leibler divergences (KL-divergence) relative to Standard Additive Model (SAM) Kernel Density Estimates (KDE) tended to converge logarithmically with increasing bandwidths. The convergence of higher-bandwidth models' log-likelihoods to the baseline's log-likelihood with itself over the interval $[0,1]$ and the

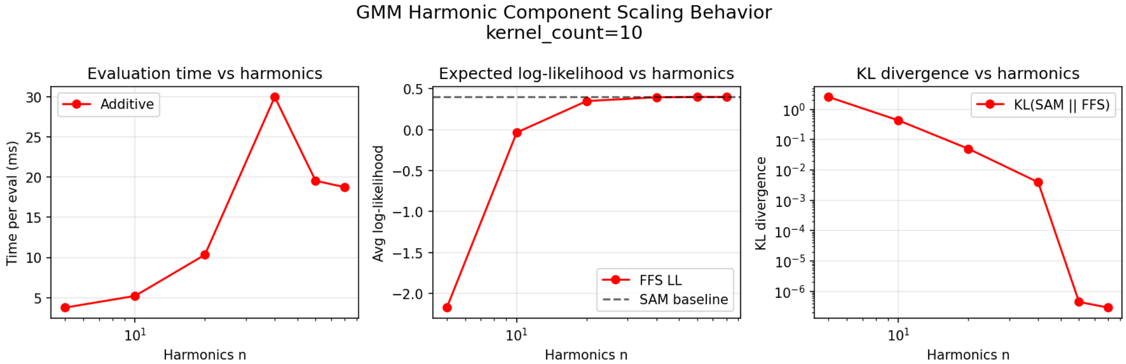

Figure 4. KDE evaluation time (left), log-likelihood of an FFS-based model with respect to a SAM Gaussian mixture (center), and KL-divergence with respect to SAM KDE (right). 10 kernels were randomly generated for each KDE's construction.

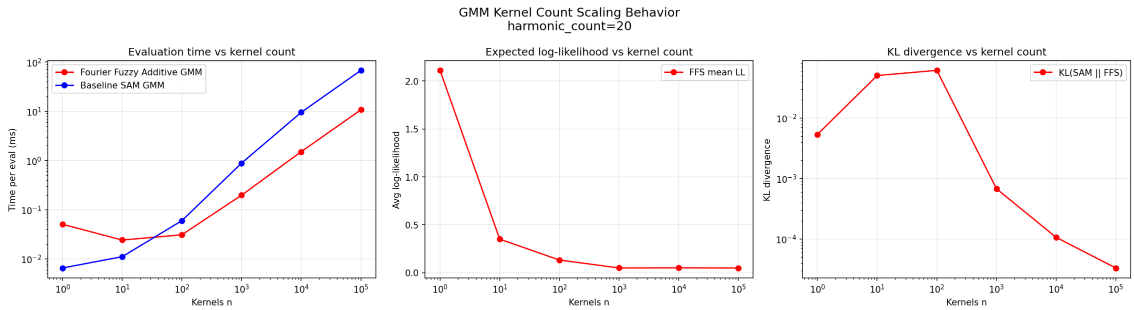

Figure 5. KDE evaluation time (left), log-likelihood of an FFS-based model with respect to a SAM Gaussian mixture (center), and KL-divergence with respect to SAM KDE (right). A fixed spectral bandwidth of 20 harmonics was used across tests.

KL-divergence rapidly converges to 0. This indicates that approximation accuracy scales well with bandwidth and our kernel approximator's limiting behavior matches the original KDE for large bandwidths.

Perhaps more interestingly, as shown in figure 5, the asymptotic FFS-based model evaluation time tended to grow at a fixed fraction of the evaluation time for the traditional SAM mixture model. For small kernel counts, it also exceeded the SAM model's evaluation time. This may be caused by hardware or implementation-specific acceleration within numpy. The rapidly diminishing log-likelihood of the FFS model relative to the SAM model may be due to the very limited spectral bandwidth used across KDEs with large kernel counts. It's probably pretty difficult to approximate a KDE with 10^5 kernels with only 20 harmonics.

Additionally, KL-divergence for large mixtures remained fairly low and beyond a peak at 100 kernels, tended to diminish to 0. This peak and subsequent convergence to 0 may be explained by the limiting behavior of large sampling distributions or visually as in figure 6.

Evaluation results are also given in tables 2 and 3 for GMMs with variable bandwidths and kernel counts respectively.

It's worth noting that much of FFS mixture models' limiting behavior presented here tends to align with that of Fourier Basis Density Models (FBMs) described in work by [4].

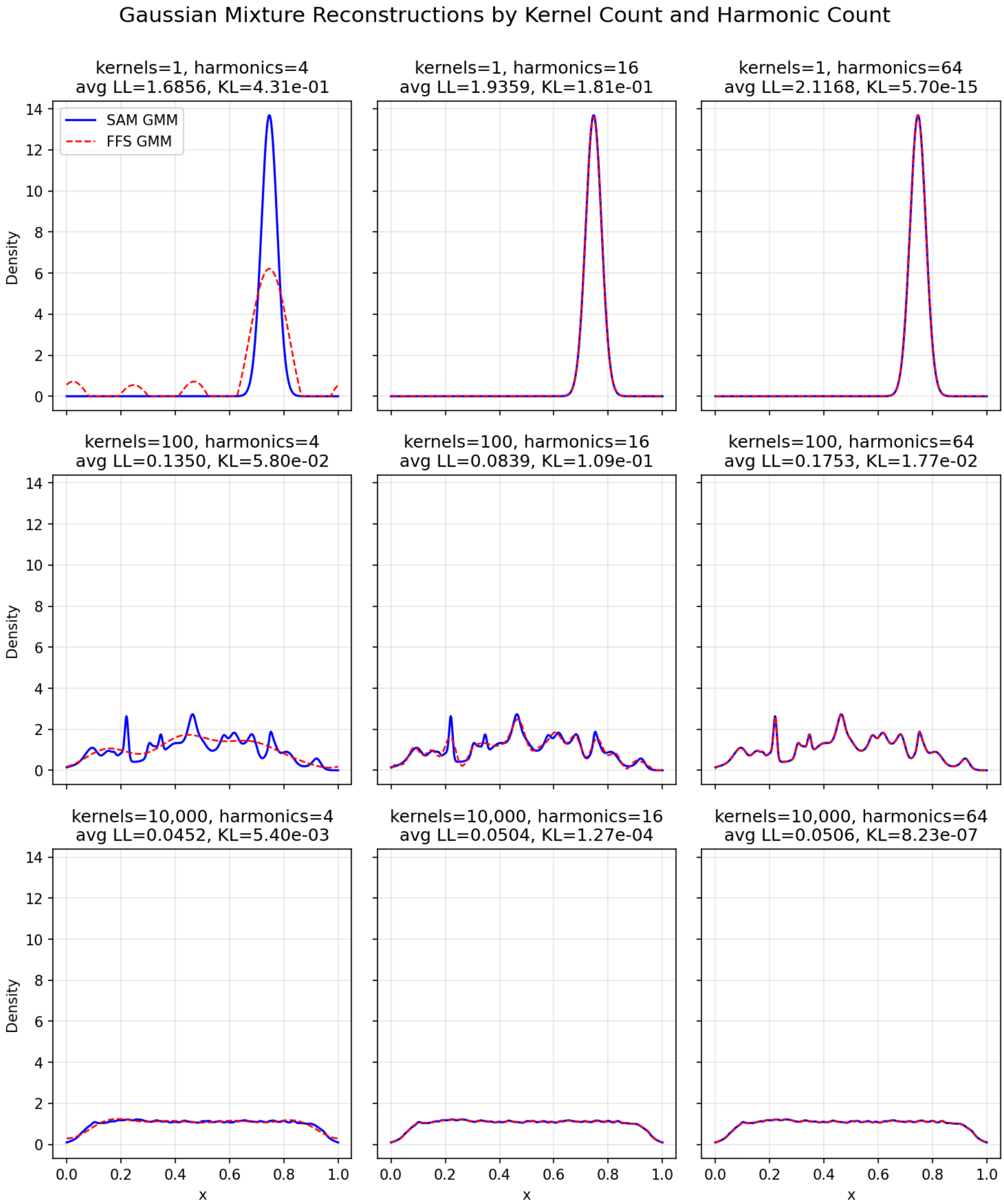

Figure 6. The result of evaluating mixtures with variable parameters. KDEs' log-likelihoods and KL-divergences tended to decrease with the number of kernels used, indicating FFS underfitting to the uniform distributions.

Number of Harmonics	FFS Time (ms)	Average Log-Likelihood	KL Divergence
5	1.98	-2.173209	2.57×10^{0}
10	2.56	-0.034542	4.35×10^{-1}
20	6.92	0.349737	5.04×10^{-2}
40	12.02	0.396212	3.95×10^{-3}
60	13.31	0.400165	4.43×10^{-7}
80	17.92	0.400165	2.89×10^{-7}

Table 2. Harmonic-count scaling results, including Fuzzy Fourier Sets' (FFSs) evaluation times. The baseline mean log-likelihood was 0.400165 and the baseline evaluation time was 0.02ms.

Number of Kernels	SAM Time (ms)	FFS Time (ms)	Avg. Log-Likelihood	KL Divergence
1	0.02	0.06	2.111393	5.37×10^{-3}
10	0.01	0.03	0.349737	5.04×10^{-2}
10^2	0.08	0.03	0.131584	6.14×10^{-2}
10^3	0.61	0.12	0.048965	6.79×10^{-4}
10^4	6.89	1.52	0.050446	1.07×10^{-4}
10^5	65.12	14.31	0.048565	3.33×10^{-5}

Table 3. Kernel-count scaling results comparing baseline additive GMM evaluation against Fourier fuzzy-set evaluation.

7. Conclusion

The development of Fourier series approximations for likelihood distribution-based fuzzy sets presents a novel and effective approach to controlling the dimensionality and interpretability of kernel density estimates. By constraining a truncated Fourier series to a domain and codomain of $[0,1]$, the composition of an arbitrary number of kernels using fuzzy logic operations with minimal impact on approximation fidelity and computational overhead. This family of fuzzy sets also demonstrates a unique robustness to vanishing limits under fuzzy logic operations through the renormalization procedure defined above. Such properties make this approach particularly well-suited for Gaussian mixture models with a large quantity of kernels, offering a competitive and scalable means for constructing complex probabilistic models.

References

[1] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965. DOI: 10.1016/S0019-9958(65)90241-X. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S001999586590241X.

[2] E. Mamdani, "Application of fuzzy algorithms for control of simple dynamic plant," *Proceedings of the Institution of Electrical Engineers*, vol. 121, pp. 1585–1588, 12 1974. DOI: 10.1049/piee.1974.0328.

[3] J. Castro, "The limits of fuzzy logic," *Mathware and soft computing, ISSN 1134-5632, Vol. 6, Nº. 2-3, 1999, pags. 155-161*, Jan. 1999.

[4] B. Kosko, "Additive fuzzy systems: From generalized mixtures to rule continua," *International Journal of Intelligent Systems*, vol. 33, no. 8, pp. 1573–1623, Jun. 2018, ISSN: 0884-8173. DOI: 10.1002/int.21925. [Online]. Available: https://doi.org/10.1002/int.21925.

[5] V. Zhurov, J. Kausch, K. Sedig, and M. Milani. "Fuzzy ontology embeddings and visual query building for ontology exploration." arXiv: 2508.08128. [Online]. Available: https://arxiv.org/abs/2508.08128.

[6] B. Okal and F. Miruka, "Application of set theory formulas in the presentation of lexical meronymy," *Saudi Journal of Humanities and Social Sciences*, vol. 1, pp. 62–72, 2016. DOI: 10.21276/sjhss.2016.1.2.5.

[7] G. Meeden, "Fuzzy set representation of a prior distribution," in *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*. Institute of Mathematical Statistics, 2008, pp. 82–88, ISBN: 9780940600751. DOI: 10.1214/074921708000000075. [Online]. Available: http://dx.doi.org/10.1214/074921708000000075.

[8] O. Osoba, S. Mitaim, and B. Kosko, "Bayesian inference with adaptive fuzzy priors and likelihoods," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2011. DOI: 10.1109/TSMCB.2011.2114879.

[9] J. Boedecker, T. Lampe, and M. Riedmiller, "Modeling effects of intrinsic and extrinsic rewards on the competition between striatal learning systems," *Frontiers in Psychology*, vol. 4, p. 739, Oct. 2013. DOI: 10.3389/fpsyg.2013.00739.

[10] Z. Ma and A. Leijon, "Bayesian estimation of beta mixture models with variational inference," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2160–2173, Nov. 2011, ISSN: 0162-8828. DOI: 10.1109/TPAMI.2011.63. [Online]. Available: https://doi.org/10.1109/TPAMI.2011.63.

[11] A. D. la Fuente, S. Singh, and J. Ballé, "Fourier basis density model," *2024 Picture Coding Symposium (PCS)*, pp. 1–5, 2024. DOI: 10.1109/PCS60826.2024.10566409. [Online]. Available: https://api.semanticscholar.org/CorpusID:267897415.

[12] D. Dubois and H. Prade, "Fuzzy numbers: An overview," in *Readings in Fuzzy Sets for Intelligent Systems*, D. Dubois, H. Prade, and R. R. Yager, Eds., Morgan Kaufmann, 1993, pp. 112–148, ISBN: 978-1-4832-1450-4. DOI: https://doi.org/10.1016/B978-1-4832-1450-4.50015-8. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9781483214504500158.

[13] S. G. Johnson. "The faddeeva package. "[Online]. Available: http://ab-initio.mit.edu/Faddeeva.