Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 9s (2025)

Probabilistic Perspective of Jungck Contractive Situations in Machine Learning

Ashutosh Mishra[1], Piyush Kumar Tripathi[2], Mukesh Kumar Singh[3], Manisha Gupta[4], Alok Kumar Agrawal[5]
[1,2,5]Department of Mathematics, ASAS, Amity University Uttar Pradesh (Lucknow Campus), India
[3]Armapore P. G. College, Kanpur, India
[4]University of Technology and Applied Science, Muscat
Email: [1]amishra94154@gmail.com, [1]ashutosh.mishra8@s.amity.edu; [2]pktripathi@amity.edu; [3]mksinghuptti@gmail.com; [4]manisha.gupta@utas.edu.om; [5]akagrawal@lko.amity.edu

Article History:

Received: 12-01-2025

Revised: 15-02-2025

Accepted: 01-03-2025

Abstract:

In the realm of machine learning (ML), the underlying mathematical frameworks significantly influence the development and efficacy of algorithms. One such framework is the probabilistic metric space, which provides a robust foundation for handling uncertainty and variability inherent in real-world problems and data. This work explores the applicability of probabilistic metric spaces in machine learning, with a particular focus on the Jungck contractive notion. The Jungck contraction theorem extends the classical Banach contraction principle to the setting of probabilistic metric spaces. This theorem is pivotal for proving the existence and uniqueness of fixed points in probabilistic settings, which is crucial for the convergence analysis of ML algorithms. This generalises invariant point propositions to probabilistic settings, offers powerful tools for the convergence analysis of iterative ML algorithms.

We delve into the theoretical underpinnings of probabilistic metric spaces, explain the Jungck contractive notion, and bespeak its relevance and application in ML algorithms.

Keywords: Probabilistic space, Contractive conditions, Machine learning, Convergence analysis.

1. Introduction

Machine learning is an underlying collection as a subset of AI (artificial intelligence) that concentrates the advancement of algorithms, as well as statistical model which enables computers to execute certain tasks without distinct instructions, depending on patterns and inference instead. In other words, it is about creating systems that can learn from data.

The crux behind such learning is to empower computer related programmes to train automatically from given data and to improve its effectiveness over time, without being distinctively programmed for every task. This is brought about through various learning techniques viz. supervised, unsupervised, reinforced, and deep.

Supervised Learning

In this approach, the algorithm retains information from some specific labelled data, meaning there by that it is given input-output pairs and learns to map inputs to outputs. For example, let us have a given set of labelled impressions (images) of cats and dogs, a supervised learning algorithm would learn to classify new impressions (images) as either cats or dogs [8].

Unsupervised Learning
In this stage, the algorithm is given input data without explicit labels, and it tries to find patterns or structure in the data. Clustering is a common task in unsupervised learning, where the algorithm groups similar data points together. An example could be grouping similar customer purchase behaviour for market segmentation [8].

Reinforced Learning
This is decision making training algorithms that gets learning from feedback. The algorithm trains itself to achieve a outcome or goal in an uncertain, potentially complex environment by taking actions and receiving rewards or penalties. It is often used in areas like gamming, robotics, and auto driving [10].

Deep Learning
It is an underlying subset of machine learning which uses artificial neural networks with many blankets of layers, hence proposed as “deep”, to model and process complex patterns in large amounts of data. It's particularly effective for tasks such as image and speech recognition, natural language processing, and playing strategic games like Go [9].

Overall, machine learning has numerous applications across various industries, including finance, e-commerce, healthcare, cyber security, and many more. Its ability to find patterns in data and make predictions or decisions based on those patterns has made it a powerful tool for solving a wide range of real-world problems.

Machine learning algorithms often rely on the convergence properties of iterative processes. Classical metric spaces and their associated invariant point theorems have provided substantial insights into these processes. However, real-world data is frequently stochastic in nature, demanding a probabilistic approach to better model the uncertainty and variability. Probabilistic metric spaces, introduced by Menger[2]in the year 1940, extend the concept of metric spaces by incorporating probability distributions instead of deterministic distances. Since then, a variety of propositions were introduced using Menger theory. Some implications of this notion are illustrated in population dynamics of cells [11, 14]. This probabilistic framework is particularly useful in ML, where data and models are inherently noisy and probabilistic.

The Jungck theorem on contraction mapping[4], a generalization of Banach's fixed point result to probabilistic metric spaces, offers a robust mechanism to analyze the convergence of sequences in these spaces[1]. This theorem is instrumental in ensuring the stability and reliability of iterative ML algorithms. By leveraging the Jungck contraction theorem, we can develop and analyze ML algorithms that are more resilient to the uncertainties of real-world data.

The *K-means* clustering algorithm is a fundamental technique in machine learning for partitioning data into clusters based on similarity[7]. In traditional *K-means*, the distance between points and centroids is deterministic. However, real-world data often contains uncertainty and noise, which can lead to instability in the clustering results. By incorporating a probabilistic approach into *K-means*, we can account for this uncertainty, leading to more robust and reliable clustering. Before delving into the probabilistic *K-means* algorithm, it's essential to understand probabilistic metric spaces (*PMS*).

2. Objectives

By employing the Jungck contractive statement, we ensure that the probabilistic gradient descent algorithm converges to an optimal solution, providing robustness against the uncertainties in the training data.

3. Methods and Preliminaries

A. Probabilistic Space: A probabilistic distance (or metric) space *(PMS)* is a generalization of a metric space where the distance between any two points is not a single real number but a distribution function. Formally, a PMS is a triplet (P, F, T), where P is a set, F is the set of distribution functions on $[0, \infty)$, and $T: P \times P \rightarrow F$is a mapping that satisfies certain axioms analogous to those of a metric space.

Definition 3.1: [11] A distribution function F on $[0, \infty)$ is a non-decreasing, left-continuous function, such that, $F(0) = 0$, and $\lim_{n \to \infty} F(x) = 1$.

The function $T(x, y)$ gives the probability that the distance between x and $y$is less than or equal to a given value. This probabilistic approach is particularly useful in modelling uncertainties in data points.

Definition 3.2: [11] A triangular norm (t-norm) can be represented as a mapping $t: [0,1] \times [0,1] \rightarrow [0,1]$ satisfying:

i. $\mathrm{t}(a, 1) = a$ for every $a \in [0,1]$
ii. $\mathrm{t}(a, b) = \mathrm{t}(b, a)$ for every $a, b \in [0,1]$
iii. $\mathrm{t}(a, c) = \mathrm{t}(b, d)$ if $a \geq b$, $c \geq d$
iv. $\mathrm{t}(a, \mathrm{t}(b, c)) = \mathrm{t}(\mathrm{t}(a, b), c); a, b, c \in [0,1]$.

B. Axioms of *PMS*:

i. Non-negativity: $T(x, y) = \epsilon_0$ (where ϵ_0 is the degenerate distribution at 0), *iff*, $x = y$.
ii. Symmetry: $T(x, y) = T(y, x), for\ all\ x \in P and\ y \in P.$
iii. Triangle inequality: For all x, y, z in P and for all $t \geq 0$,

$$T(x, z)(t) \geq \sup_{u+v=t} min\, \{T(x, y)(u), T(y, z)(v)\}.$$

These axioms ensure that the probabilistic metric retains the essential properties of a classical metric, allowing us to extend various analytical techniques to the probabilistic domain.

***C.* Jungck Contractive Condition:** The Jungck contraction theorem extends the classical Banach contraction principle to the setting of probabilistic metric spaces. This theorem is pivotal for proving the existence and uniqueness of fixed points in probabilistic settings, which is crucial for the convergence analysis of ML algorithms.

Theorem 3.1 (Jungck Contraction Theorem): [4] Let (P, F, T) be a complete probabilistic metric space. Suppose $f, g: P \rightarrow P$are two mappings satisfying:

$$T(f(x), f(y))(t) \geq \beta\, T(g(x), g(y))(t),$$

for all $x, y \in P$ and for some constant $\beta \in (0,1)$.Then,f and g have a unique common invariant point in P.

This result ensures that under a contraction condition, iterative applications of the mappings f and g will converge to a common invariant point. This result is useful in analyzing machine learning based algorithms where such mappings often represent iterative update steps, for example, in analysing the convergence of probabilistic *K-means* clustering.

Application in Machine Learning: Many algorithms, in machine learning, can be viewed as iterative processes that seek to minimize an objective function or optimize a model. The convergence of these processes is crucial for the reliability and stability of the algorithms. By framing these iterative processes within the context of probabilistic metric spaces, we can leverage the Jungck contractive condition to assure convergence even in the presence of uncertainty.

***K-means* Clustering in Probabilistic Space-A Case Study:** The *K-means* clustering algorithm is a widely used method for partitioning data into clusters. However, in the presence of uncertainty and noise, the classical *K-means* algorithm may struggle to find stable clusters. By adopting a probabilistic approach, we can enhance the robustness of *K-means* clustering.

Probabilistic Distance Metric: Instead of using a deterministic distance, we define a probabilistic distance between points. This distance is modelled as a distribution function capturing the uncertainty in the data.

Probabilistic Centroid Update: The update step for centroids in the classical *K-means* is replaced by a probabilistic update. Given a cluster, the new centroid is chosen such that it minimizes the expected probabilistic distance to the points in the cluster.

Convergence Analysis: Using the Jungck contraction theorem, we can analyse the convergence of the probabilistic *K-means* algorithm. By ensuring that the probabilistic distance satisfies the contraction condition, we can guarantee that the iterative updates of centroids will converge to a stable configuration.

***K-means* Clustering Algorithm in Probabilistic Space:** The probabilistic *K-means* algorithm [7] extends the classical *K-means* by incorporating probabilistic distances, which are represented by distribution functions. This approach provides a more robust clustering method in the presence of uncertainty and noise.

Algorithm of *K-means* in probabilistic space is as follows:

1. Initialization: Randomly initialize k probabilistic centroids, where each centroid is represented by a distribution function.

2. Assignment Step: Here, we assign each data point to the cluster whose centroid minimizes the probabilistic distance.

3. Update Step: For each and every cluster, update the centroid to minimize the expected probabilistic distance to the points in the cluster.

4. Convergence Check: Repeat the assignment and update steps until the centroids converge.

Step-by-Step Explanation:

1. Initialization:

Randomly select k points from the dataset as initial centroids.

Represent each centroid as a distribution function capturing the uncertainty.

2. Assignment Step:

For each data point x_i, compute the probabilistic distance to each centroid C_j using the distribution function $T(x_i, C_j)$.

Assign $x_i$to the cluster with the centroid that has the smallest expected probabilistic distance.

3. Update Step:

For each cluster, C_j, update the centroid to the point that minimizes the expected probabilistic distance to all points in the cluster.

This involves finding a new distribution function that best represents the central tendency of the points in the cluster.

4. Convergence Check:

We check, if the centroids have stabilized (i.e., there is little to no change in the distribution functions of the centroids).

If not so, we repeat the assignment and update steps until the convergence is obtained, e.g.

Consider a simple 2D dataset with three clusters. The classical *K-means* algorithm may struggle to identify the clusters accurately due to noise and uncertainty in the data points. We apply the probabilistic *K-means* algorithm to handle this uncertainty.

Dataset:

Points: [(1.0, 2.0), (1.5, 1.8), (5.0, 8.0), (8.0, 8.0), (1.0, 0.6), (9.0, 11.0)]

Initialization:

Randomly initialize $k = 2$ probabilistic centroids. Assume initial centroids are $(1.0, 2.0)$ and $(5.0, 8.0)$, represented by distribution functions with small variances.

Iteration 1: Assignment Step

Compute probabilistic distances between each point and the centroids using a Gaussian distribution function.

Assign points to the nearest centroid based on expected probabilistic distance.

Iteration 1: Update Step

Update centroids by recalculating the distribution functions that minimize the expected probabilistic distance within each cluster.

Iteration 1: Results

Cluster 1: [(1.0, 2.0), (1.5, 1.8), (1.0, 0.6)]

Cluster 2: [(5.0, 8.0), (8.0, 8.0), (9.0, 11.0)]

New Centroids (distribution functions):

Cluster 1: Mean = (1.17, 1.47), Variance = small

Cluster 2: Mean = (7.33, 9.0), Variance = small

Iteration 2: Assignment Step

Recompute probabilistic distances using updated centroids.

Reassign points to the nearest centroid.

Iteration 2: Update Step

Recalculate centroids for the new clusters.

Iteration 2: Results

Cluster 1: [(1.0, 2.0), (1.5, 1.8), (1.0, 0.6)]

Cluster 2: [(5.0, 8.0), (8.0, 8.0), (9.0, 11.0)]

New Centroids (distribution functions):

Cluster 1: Mean = (1.17, 1.47), Variance = very small

Cluster 2: Mean = (7.33, 9.0), Variance = very small

Convergence Check:

Centroids have stabilized (no significant change in mean and variance of distribution functions).

Algorithm converges.

Final Clusters:

Cluster 1: [(1.0, 2.0), (1.5, 1.8), (1.0, 0.6)]

Cluster 2: [(5.0, 8.0), (8.0, 8.0), (9.0, 11.0)]

By applying the Jungck contraction theorem, we ensure that the probabilistic *K-means* algorithm converges to a stable clustering configuration, providing robustness against the inherent uncertainties in the data.

Probabilistic Neural Networks (PNN)-A Case Study: Neural networks (NNs) are a cornerstone of modern ML. Training a NN involves iteratively updating weights to minimize a loss function. In the presence of noisy data, a probabilistic framework can enhance the training process[13].PNNs and related probabilistic models are being applied to various domains such as healthcare (medical diagnosis and prognosis) and finance (risk assessment and fraud detection). These applications benefit from the ability of PNNs to handle uncertainty and complex patterns in data[13].

1. Probabilistic Weight Updates: The weight updates are modelled probabilistically, capturing the uncertainty in the gradient estimates due to noisy data.

2. Probabilistic Loss Function: The loss function is defined in terms of probabilistic distances, providing a more robust objective for optimization [16].

3. Convergence Analysis: Using the Jungck contraction theorem, we analyse the convergence of the probabilistic weight updates [16]. By ensuring that the probabilistic distance satisfies the contraction condition, we can assure convergence to an optimal set of weights.

Probabilistic Gradient Descent Algorithm:

1. Initialization: Randomly initialize weights probabilistically.

2. Forward Pass: Compute the output of the NN using the current probabilistic weights.

3. Loss Computation: Compute the probabilistic loss based on the output and the true labels.

4. Backward Pass: Compute the probabilistic gradient of the loss with respect to the weights.

5. Weight Update: Update the weights probabilistically based on the computed gradient.

6. Convergence Check: We repeat steps 2-5, until the loss converges.

By employing the Jungck contractive statement and commutativity [15], we ensure that the probabilistic gradient descent algorithm converges to an optimal solution, providing robustness against the uncertainties in the training data.

4. Main Results & Statements

We use the following as a result in harnessing the situation as stated above.

Lemma 4.1: [6] If (P, F, T) be a Menger generalized PM-space with a Cauchy sequence $< \alpha_n >$ in Palong with some $\beta \in (0,1)$ such that $F_{\alpha_n,\alpha_{n+1}}(\beta x) \geq F_{\alpha_{n-1},\alpha_n}(x)$ for all positive values of x.

Also, if for $r > 1$ and $i = n\ to\ \infty$, we have

$$\lim_{n\to\infty} T_i\left(F_{\alpha_0,\alpha_1}(r^i)\right) = 1$$

Therefore, we say the sequence $< \alpha_n >$ is F-Cauchy.

To get the common coincidence point for mappings under generalized Menger space for producing some machine learning results, we can apply the following statements with some specific constraints. Theorem 4.1: With (P, F, T)a generalized PM-space and a continuous t-norm$T$ in $(a, 1)$ for all $0 < a < 1$; $0 < \theta < 1$; $\delta \in \Delta$ and $f, g: Q \to P$, we have

(i). $\delta(F_{f\alpha,f\beta}(\theta x), F_{g\alpha,g\beta}(x), F_{f\alpha,g\alpha}(x), F_{f\beta,g\beta}(\theta x)) \geq 0$ for every $\alpha, \beta \in Q$ and positive values of x
(ii).$f(Q) \subset g(Q)$
(iii).$\alpha_0, \alpha_1 \in Q$ so that $f\alpha_0 = g\alpha_1$ and $\lim_{n\to\infty} T_i\left(F_{f\alpha_0,f\alpha_1}(r^i)\right) = 1$for $r > 1$ and $i = n\ to\ \infty$
Then, we can say to have common coincidence value for$f$and g.

Proof: Suppose, $\alpha_0, \alpha_1 \in Q$ so that $f\alpha_0 = g\alpha_1$ and $\lim_{n\to\infty} T_i\left(F_{f\alpha_0,f\alpha_1}(r^i)\right) = 1$ (1)

As $f(Q) \subset g(Q)$ and $f\alpha_0 = g\alpha_1$, so a sequence $< \alpha_n >$ can be constructed to have $f\alpha_n = g\alpha_{n+1}$

Consider,$z_n = f\alpha_n$

Then, (1) gives, $\lim_{n\to\infty} T_i\left(F_{z_0,z_1}(r^i)\right) = 1$; $i = n\ to\ \infty$. Substituting $\alpha = \alpha_n$ and $\beta = \beta_{n+1}$ in (i), we get

$$\delta(F_{f\alpha_n,f\alpha_{n+1}}(\theta x), F_{g\alpha_n,g\alpha_{n+1}}(x), F_{f\alpha_n,g\alpha_n}(x), F_{f\alpha_{n+1},g\alpha_{n+1}}(\theta x)) \geq 0,$$

$$i.e.\ \ \delta(F_{z_n,z_{n+1}}(\theta x), F_{z_{n-1},z_n}(x), F_{z_{n-1},z_n}(x), F_{z_n,z_{n+1}}(\theta x)) \geq 0$$

and using virtue of δ, for all n, we obtain

$$F_{z_n,z_{n+1}}(\theta x) \geq F_{z_{n-1},z_n}(x)$$

So, by above Lemma, the sequence $< z_n >=< f\alpha_n >$ is Cauchy.

If $g(Q)$ is F- complete, then there must be $\alpha \in g(Q)$, so that,$< z_n > \to \alpha$, and for any $z \in Q$ so that $gz = \alpha$. Substituting $\alpha = \alpha_n$ and $y = z$, in (i), we have

$$\delta(F_{f\alpha_n,fz}(\theta x), F_{g\alpha_n,gz}(x), F_{f\alpha_n,g\alpha_n}(x), F_{fz,gz}(\theta x)) \geq 0$$

$$i.e.\ \delta(F_{z_n,fz}(\theta x), F_{z_{n-1},gz}(x), F_{z_{n-1},fz}(x), F_{gz,fz}(\theta x)) \geq 0$$

As $n \to \infty$, we get,

$$\delta(F_{gz,fz}(\theta x), F_{gz,gz}(x), F_{gz,fz}(x), F_{fz,gz}(\theta x)) \geq 0$$

$$i.e.\ \delta(F_{gz,fz}(\theta x), 1, F_{gz,fz}(x), F_{fz,gz}(\theta x)) \geq 0$$

Again, as $F_{gz,fz}(\theta x) \geq 1$, so $fz = gz = \alpha$

Taking limit as $n \to \infty$, we have,

Hence, f and g have$z$ as a coincidence point and thus, for $f(Q)$ to be complete, we have $< z_n > \to \alpha \in f(Q) \subset g(Q)$ so, z is a common coincidence point.

Theorem 4.2: With (P, F, T)a generalized PM-space and a continuous t-norm$T$ in $(a, 1)$ for all $0 < a < 1$; $0 < \theta < 1$; $\delta \in \Delta$ and $f, g: Q \to P$, we have

(i). $\delta(F_{f\alpha,f\beta}(\theta x), F_{g\alpha,g\beta}(x), F_{f\alpha,g\alpha}(x), F_{f\beta,g\beta}(\theta x)) \geq 0$ for every $\alpha, \beta \in Q$ and positive values of x
(ii).$f(P) \subset g(P)$
(iii). $\exists\ \alpha_0, \alpha_1$ so that $f\alpha_0 = g\alpha_1$ and $\lim_{n\to\infty} T_i\left(F_{f\alpha_0,f\alpha_1}(r^i)\right) = 1$for $r > 1$ and $i = n\ to\ \infty$
(iv). Either $f(P)$ or $g(P)$ is F- complete
(v). f and g are commuting functions that commute at their coincidence point.
Then, f and g have a common invariant point that is unique in, P.

Proof: Let us choose $P = Q$ in previous theorem, then we obtain $z_n = f\alpha_n$ and sequence $< z_n >$ is Cauchy. If $g(P)$ is F- complete, then $z_n \to \alpha \in g(P)$, and we have $z \in P$ so that $g(z) = \alpha$. Replacing α by α_n and β by z in inequality(i) we have; $fz = gz = \alpha$. As f and g are commuting, therefore, we must have, $fgz = gfz$, so that,$f\alpha = g\alpha$.

Again, using $\alpha = z$ and $\beta = fz$ in inequality (i), we get

$$\delta(F_{fz,ffz}(\theta x), F_{gz,gfz}(x), F_{gz,fz}(x), F_{ffz,gfz}(\theta x)) \geq 0$$

$$i.e.\,\delta(F_{\alpha,f\alpha}(\theta x), F_{\alpha,f\alpha}(x), 1, F_{f\alpha,g\alpha}(\theta x)) \geq 0$$

Using the properties of distance metric(δ) for all positive,x we obtain, $F_{\alpha,f\alpha}(\theta x) \geq F_{\alpha,f\alpha}(x)$ which is true only if $\alpha = f\alpha$, otherwise we reach to a contradiction with respect to Lemma 4.1. This implies that, α is a common invariant point of f and g.

To prove the uniqueness, we suppose α and β are two common invariant points of f and g.

Using (i), we get

$$\delta(F_{\alpha,\beta}(\theta x), F_{\alpha,\beta}(x), F_{\alpha,\alpha}(x), F_{\beta,\beta}(\theta x)) \geq 0 \Rightarrow F_{\alpha,\beta}(\theta x) \geq 1 \Rightarrow \alpha = \beta.$$

This shows the uniqueness of the result.

Definition 4.1: [6] Aforementioned t–norm T is said Hadžić type or H- type, denoted as $T \in H$, if the group $\{T^n\}_{n\in N}$ of it iterates is inductively depicted, $\forall x \in [0,1]$ by,

$T^0(x) = 1, T^{n+1}(x) = T(T^n(x), x)$, for all non-negative x and as T^n is defined, it is equi-continuous at $x = 1$. This states the following,

$$\varepsilon \in (0,1) \exists \text{ a } \vartheta \in (0,1), \text{ so that, } x > 1 - \vartheta \Rightarrow T^n(x) > 1 - \varepsilon, \text{ for all } n \geq 1.$$

The statement is followed by Hadžić:

Statement 4.1: [6]

(i). If we take $T \geq T_L$, then for $i = n\ to\ \infty$,

$$\lim_{n\to\infty} T_i x_{n+1} = 1, \text{ both sides imply } \sum 1 - x_n < \infty \text{ and } T = T_\mu^{SW}$$

$$(\text{By } T \geq T_L\ i.e.\, a \geq c, b \geq d \Rightarrow T(a,b) = T(c,d)).$$

(ii).When, $T \in H$, then for each $(x_n)_{n\in N}$ in [0,1],$\lim_{n\to\infty} x_n = 1 \Rightarrow \lim_{n\to\infty} T_i x_{n+i} = 1$,for $i = n\ to\ \infty$.

(iii). When, $T \in T_\mu^D, T_\mu^{AA} T$, then for $i = n\ to\ \infty$, $\lim_{n\to\infty} T_i x_{n+i} = 1$, both sides imply $\sum(1 - x_n)^\mu < \infty$.

We wish to repeat the notion of distribution functions and PMS for the interval$[0,\infty]$.

Definition 4.2: [6] Consider a class of distribution mapL_+as $F: [0,\infty] \to [0,\infty]$ with,

(i). F has zero value at $x = 0$
(ii).F is a non-decreasing map and
(iii). F is left continuous for $0 < x < \infty$.
Let $D_+ \subset L_+$ contains$F$, so that, $\lim_{n\to\infty} F(x) = 1$. Then, the distribution function, ε_0 is defined as ε_0

$$\varepsilon_0 = \begin{cases} 0, & x = 0 \\ 1, & x > 0 \end{cases} \text{ lies in } D_+.$$

Statement 4.2: Consider (P, F, T) asa generalized PM-space and a continuous t-norm$T \in H$ with $0 < \theta < 1$; $\delta \in \Delta$ and $f, g: P \to P$, so that,

(i). $\delta(F_{f\alpha,f\beta}(\theta x), F_{g\alpha,g\beta}(x), F_{f\alpha,g\alpha}(x), F_{f\beta,g\beta}(\theta x)) \geq 0$ for every $\alpha, \beta \in P$ and positive values of x
(ii). $f(P) \subset g(P)$
(iii). $\exists\ \alpha_0, \alpha_1$ so that $f\alpha_0 = g\alpha_1$ for which $F_{f\alpha_0,f\alpha_1} \in D_+$
(iv). Either $f(P)$ or $g(P)$ is F- complete

(vi). f and g are commuting functions that commute at their coincidence point.
Then, f and g have a unique common invariant point in P.

Proof: In Theorem 4.1, if we consider a $\mu > 1$ with a sequence $< z_n >$ along with a mapping, $f_{z_0,z_1} \in D_+$, so that, $\lim_{n\to\infty} F_{z_0,z_1}(\mu^n) = 1$. Therefore, using (ii) of Statement 4.1, we get

$$\lim_{n\to\infty} T_i\left(F_{z_0,z_1}(\mu^i)\right) = 1; i = n\ to\ \infty.$$

Hence, Theorem 4.2 follows the required results.

5. Discussion

Probabilistic spaces offer a powerful framework for handling the uncertainties and variability inherent in real-world data [12]. By incorporating probability distributions into the metric space, we can model the stochastic nature of data more effectively. The Jungck contraction theorem provides a robust tool for analysing the convergence of iterative processes in probabilistic metric spaces, which is crucial for the stability and reliability of ML algorithms. This work has demonstrated the application of probabilistic metric spaces and the Jungck contraction theorem in enhancing the robustness and convergence properties of ML algorithms. Through case studies on probabilistic *K-means* clustering and probabilistic neural networks, we have illustrated the practical benefits of this approach. Future research can extend these concepts to other ML algorithms and explore the integration of probabilistic metric spaces with other probabilistic modelling techniques.

In a nutshell, the probabilistic metric space approach, underpinned by the Jungck contraction theorem, offers a promising direction for developing more robust and reliable ML algorithms in the face of uncertainty and variability.

References

[1] S. Banach, “Sur les opérations dans les ensembles abstraits et leur application aux équationsintégrales,” *Fundamenta Mathematicae*, 3(1), pp.133-181,1922.

[2] K. Menger, “Statistical metrics,” *Proceedings of the National Academy of Sciences of the United States of America*, 28(12), pp.535-537, 1942.

[3] E. Fix, and J. L. Hodges, “Discriminatory analysis-nonparametric discrimination: Consistency properties. Report Number 4, Project 21-49-004,” *USAF School of Aviation Medicine*, Randolph Field, Texas,1951.

[4] G. Jungck, “Compatible mappings and common fixed points,” International *Journal of Mathematics and Mathematical Sciences*, 1(4), pp.215-222,1976.

[5] C. C. Y. Dorea, “On probabilistic metric spaces,” *Bulletin of the Brazilian Mathematical Society*, 14(2), pp.99-118,1983.

[6] O. Hadžić and E. Pap, “Probabilistic metric spaces. In: Fixed point theory in probabilistic metric spaces,” *Mathematics and its Appl*. Vol. 536, Springer, Dordrecht, pp.47-94, 2001.https://doi.org/10.1007/978-94-017-1560-7_2

[7] O. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, "Clustering with Bregman Divergences." *Journal of Machine Learning Research*, vol. 6, pp. 1705-1749, 2005.

[8] T. Hastie, R. Tibshirani, J. Friedman, “The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.),” Springer, 2009.

[9] I. Goodfellow, Y. Bengio, & A. Courville, “Deep Learning,” *MIT Press*, 2016.

[10] R. S. Suttonand A. G. Barto, “Reinforcement Learning: An Introduction (2nd ed.)”. *MIT Press*, 2018.

[11] A. Mishra, P. K. Tripathi, A. K. Agrawal, D. R. Joshi, “A Probabilistic Approach in Modeling the Growth of Cell Population,” *Test Engineering and Management*, Vol. 83, pp. 737-742, 2020.

[12] A. Mishra, Achieving Environmental Sustainability Through Contraction Mapping in Growth of Cell Population, *International Journal of Engineering Research and Management Technology*, ISSN:2348-403, International Conference on Mathematical Sciences and Computational Intelligence (ICMSCI-2020), (2020), 35-36.

[13] V. Fortuin, O. E. Ganeaand M. Welling, “Deep generative models with learnable knowledge priors,” *In International Conference on Machine Learning (ICML)*, 2021.

[14] A. Mishra and P. K. Tripathi, “Iteration-based reduction in cell population for biomedical applications”, *Proceedings of Advances in AI for Biomedical Instrumentation, Electronics and Computing*, Taylor and Francis Group, CRC Press, UK, pp.466-72, 2024. http://dx.doi.org/10.1201/9781032644752-86

[15] A. Mishra and P. K. Tripathi, “Applicability of Intimate Mapping in Digital Image Sources”, *Grenze International Journal of Engineering and Technology* (*Grenze Scientific Society*), 10 (1), 711-715, 2024.

[16] A. Mishra and P. K. Tripathi, “Iterative Methods under Jungck Contraction Theorem for Root Finding in Numerical Analysis for AI Applications”, *Grenze International Journal of Engineering and Technology* (*Grenze Scientific Society*), 11 (1), 4703-4708, 2025.