ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 5, No. 1, 2010, *pp*. 019-040

Structural Risk Minimization Principle Based on Complex Fuzzy Random Samples*

Zhiming Zhang [+a,]*, Jingfeng Tian[b]

[a]College of Mathematics and Computer Sciences, Hebei University, 180 Wu Si East Road, Baoding 071002, Hebei Province, PR China

[b]Science and Technology College, North China Electric Power University, 282 Rui Xiang Street, Baoding 071051, Hebei Province, PR China

(*Received February 30, 2009, accepted April 5, 2009*)

Abstract. Statistical Learning Theory is commonly regarded as a sound framework within which we handle a variety of learning problems in presence of small size data samples. It has become a rapidly progressing research area in machine learning. The theory is based on real random samples and as such is not ready to deal with the statistical learning problems involving complex fuzzy random samples, which we may encounter in real world scenarios. This paper explores statistical learning theory based on complex fuzzy random samples. Firstly, the definition of complex fuzzy random variable is introduced. Next the concepts and some properties of the mathematical expectation and independence of complex fuzzy random variables are provided. Secondly, the concepts of annealed entropy, growth function and VC dimension of measurable complex fuzzy set valued functions are proposed, and the bounds on the rate of uniform convergence of learning process based on complex fuzzy random samples are constructed. Thirdly, on the basis of these bounds, the idea of the complex fuzzy structural risk minimization principle is presented. Finally, the consistency of this principle is proven and the bound on the asymptotic rate of convergence is derived.

Keywords: complex fuzzy random variable, annealed entropy, growth function, VC dimension, complex fuzzy structural risk minimization principle, bound on the asymptotic rate of convergence

1. Introduction

Statistical Learning Theory (SLT, for short), proposed in 1960s and fully established in 1990s by Vapnik et al. [27-29], has emerged as an interesting and sound theory that supports the development of laws of statistical learning for small data samples. It has provided effective solutions obtained in presence of small samples where such samples are inherently associated with crucial issues such as overfitting and underfitting, high-dimensionality of classification problems, existence of multiplicity of local minima and other important problems encountered in practice of machine learning methods and their architectures such as e.g., neural networks. In the late 1990s SLT had become one of the fastest-growing disciplines in machine learning. Its essence was to make the learning machines work effectively with the limited samples and then improve the generalization abilities of the learning machines. By doing this, we establish a meaningful theoretical framework for statistical learning for small data samples. Meanwhile, SLT gave rise to a new category of general learning algorithms, namely Support Vector Machine (SVM, for short). At present, the SLT and SVM constitute interesting research avenues in machine learning [1,5-7,12,13,15,24,26,30,32,34].

Despite the fact that SLT has reached a substantial level of maturity, there are still a number of open issue as e.g., the development of the SLT and SVM realized on a basis of probability measure space and the real-valued random samples (real numbers-valued random variables). In real world scenarios, there often are many non-probability spaces (such as fuzzy measure spaces [33], credibility measure space [19], etc.) and non-real valued random samples (such as fuzzy random samples [16], complex random samples [36], etc.).

* This work was supported by the National Natural Science Foundation of China (Grant No. 60773062), the Natural Science Foundation of Hebei Province of China (Grant No. 2008000633), a Key Scientific Research Project of Department of Education of Hebei Province of China (Grant No. 2005001D), and the Key Scientific and Technical Research Project of Ministry of Education of China (Grant No. 206012).

+ Corresponding author. *E-mail address:* zhangzm1978@yahoo.com.cn.

To alleviate these problems, it becomes imperative to move forward with extensions and generalizations along the line of dealing with the statistical learning theories established on non-probability space and based on non-real valued random samples. Some research [8-11,14,18,20,25] has been already realized along this line. For example, Ha et al. [8-10] generalized the key theorem of learning theory and constructed the bounds on the rate of convergence of learning process of statistical learning theory form probability space to Sugeno measure space, credibility space and quasi-probability spaces where these three are typical non-probability spaces; Lin and Wang [18] constructed support vector machine based on fuzzy random samples; Liu and Chen [20] discussed face recognition using total margin-based adaptive fuzzy support vector machines; Jin, Tang and Zhang [14] constructed support vector machines with genetic fuzzy feature transformation for biomedical data classification.

However, there has been rather little work completed for statistical learning theory based on complex fuzzy random samples. It is well known that complex numbers constitute a substantial and practically relevant generalization of real numbers. By the same token, samples formed by complex fuzzy random variables constitute the important generalization of real random variables. The work on fuzzy complex analysis was started by the concept of fuzzy complex numbers which introduced at first by Buckley [2]. In the sequel, there have been further researches on fuzzy complex analysis [3,4,21-23,31,35,37,38]. For example, Buckley [3,4] discussed the differentiability and integrability of fuzzy complex valued functions; Zhang [37] presented the limit theory of the sequence of fuzzy complex numbers, and given a series of results about limit theory; Qiu, Wu and Li [22,23] revisited the idea of fuzzy complex analysis in several different ways. They looked at different concepts of convergence and relationships between them and discussed the continuity and differentiation of fuzzy complex functions. Zhang [38] proposed the concepts of measurable complex fuzzy set valued function and complex random variable, and discussed some properties in detail. All these observations lead to the conclusion of relevance and applicability of the generalization of the SLT to complex fuzzy random samples (we note that the fuzzy random samples and the complex random samples are two special cases). In the SLT, one of the kernels content is a new induction principle, the so-called "structural risk minimization" principle, which is a better induction principle of the learning machine than the empirical risk minimization principle. This principle minimizes bounds with respect to two factors, the value of empirical risk and the capacity. Moreover this principle allows us to find the function that achieves the guaranteed minimum of the expected risk using a finite number of observations. However, in the classical statistical learning theory, the conclusions about the structural risk minimization principle were based on real random samples, which are not ready to deal with the statistical learning problems involving complex fuzzy random samples which we may encounter in real world scenarios. This study first proposes the concept of structural risk minimization principle of complex fuzzy random samples by combining fuzzy complex analysis and SLT, then the consistency of the structural risk minimization principle of complex fuzzy random samples and asymptotic bounds on the rate of convergence are presented and proven. The study will help lay essential theoretical foundations for support vector machine based on complex fuzzy random samples.

This paper is organized as follows. Section 2 introduces some basic definitions and properties which will be used in the study. In Section 3, the concepts of capacity for the set of measurable complex fuzzy set valued functions are proposed. In the sequel, in Section 4, we give the bounds on the rate of uniform convergence of learning process based on complex fuzzy random samples. In Section 5, we propose the concepts of complex fuzzy structural risk minimization principle. In Section 6, we prove the consistency of the complex fuzzy structural risk minimization principle and construct asymptotic bound on the rate of convergence. The final section offers the conclusions and brings prospects of potential future developments.

2. Preliminaries

Throughout this paper, we assume that $(\Omega, \mathcal{A}, P)$ is a probability measure space, $\mathbb{R}$ is the real numbers field, and $\mathcal{K}_c(\mathbb{R})$ is a family of nonempty compact convex subsets of $\mathbb{R}$. Let $\mathcal{F}(\mathbb{R})$ denote the family of all functions $\tilde{X}:\mathbb{R}\to[0,1]$, and Let $\mathcal{F}_c(\mathbb{R})$ denote the family of all functions $\tilde{X}\in\mathcal{F}(\mathbb{R})$ which satisfies the following conditions:

(1) $\tilde{X}$ is normal, i.e., there exists $x\in\mathbb{R}$ such that $\tilde{X}(x)=1$;

(2) $\tilde{X}$ is upper semi-continuous;

(3) $\text{supp}\,\tilde{X} = cl\left\{x \in \mathbb{R} : \tilde{X}(x) > 0\right\}$ is compact;

(4) $\tilde{X}$ is a convex fuzzy set, i.e., $\tilde{X}(\lambda x + (1-\lambda)y) \geq \min(\tilde{X}(x), \tilde{X}(y))$, for $\forall x, y \in \mathbb{R}$ and $\lambda \in [0,1]$.

For a fuzzy set $\tilde{X} \in \mathcal{F}(\mathbb{R})$, if we define

$$(\tilde{X})_r = X_r = \begin{cases} \left\{x : \tilde{X}(x) \geq r\right\}, 0 < r \leq 1 \\ \text{supp}\,\tilde{X}, r = 0 \end{cases},$$

then it follows that $\tilde{X} \in \mathcal{F}_c(\mathbb{R})$ if and only if $X_1 \neq \Phi$ and X_r is a closed bounded interval for each $r \in [0,1]$. Therefore, $\tilde{X}$ is completely determined by the interval $X_r = \left[X_r^-, X_r^+\right]$.

If $A, B \in \mathcal{K}_c(\mathbb{R})$, then the Hausdorff metric is defined by

$$d_H(A,B) = \max\left\{\sup_{x\in A}\inf_{y\in B}|x-y|, \sup_{y\in B}\inf_{x\in A}|x-y|\right\}.$$

Let $\tilde{X}, \tilde{Y} \in \mathcal{F}_c(\mathbb{R})$, and set

$$d(\tilde{X}, \tilde{Y}) = \sup_{0<r\leq 1} d_H(X_r, Y_r),$$

where $d_H(X_r, Y_r) = \max\left\{\left|X_r^- - Y_r^-\right|, \left|X_r^+ - Y_r^+\right|\right\}$. Also, the norm $\|\tilde{X}\|$ of fuzzy number $\tilde{X}$ will be defined as

$$\|\tilde{X}\| = d(\tilde{X}, \tilde{0}) = \max\left\{\left|X_0^-\right|, \left|X_0^+\right|\right\}.$$

Definition 2.1 [38]**.** Let $\mathbb{C}$ be a complex numbers field. The mapping $\tilde{Z} : \mathbb{C} \to [0,1]$ is called a fuzzy complex set. The $r-$cut of $\tilde{Z}$ is

$$(\tilde{Z})_r = Z_r = \left\{z \middle| \tilde{Z}(z) \geq r, r \in (0,1]\right\}.$$

We separately define Z_0, the $0-$cut of $\tilde{Z}$, as the closure of the union of the Z_r for $0 < r \leq 1$.

Definition 2.2 [38]**.** If $\forall \tilde{X}, \tilde{Y} \in \mathcal{F}(\mathbb{R})$ with membership functions $\mu(x|\tilde{X})$ and $\mu(y|\tilde{Y})$, respectively, then $\tilde{Z} = \tilde{X} + i\tilde{Y}$ is called a complex fuzzy set with membership function $\mu(z|\tilde{Z}) = \min\left\{\mu(x|\tilde{X}), \mu(y|\tilde{Y})\right\}$, where $z = x + iy$. We denote the class of all the complex fuzzy sets by $\mathcal{F}(\mathbb{C})$. Especially, if $\tilde{X}, \tilde{Y} \in \mathcal{F}_c(\mathbb{R})$, then we call $\tilde{Z} = \tilde{X} + i\tilde{Y}$ a bounded closed complex fuzzy number. Because $(\tilde{Z})_r = X_r \times Y_r = X_r + iY_r$ for $0 \leq r \leq 1$ and the r-cuts of $\tilde{Z}$ are rectangles, a bounded closed complex fuzzy number is also called a rectangular fuzzy complex number (see Ref. [2]). And we denote by $\mathcal{F}_c(\mathbb{C})$ the class of all the bounded closed complex fuzzy numbers, i.e., $\mathcal{F}_c(\mathbb{C}) = \left\{\tilde{Z} = \tilde{X} + i\tilde{Y} \middle| \tilde{X}, \tilde{Y} \in \mathcal{F}_c(\mathbb{R})\right\}$.

Let $\mathcal{K}_c(\mathbb{C})$ be a family of nonempty compact convex subsets of $\mathbb{C}$. If $A, B \in \mathcal{K}_c(\mathbb{C})$, then the Hausdorff metric is defined by

$$d_H(A,B) = \max\left\{\sup_{x\in A}\inf_{y\in B} d(x,y), \sup_{y\in B}\inf_{x\in A} d(x,y)\right\},$$

where $d(x,y)$ denotes the distance between two complex numbers x and y.

Let us define a consistent Hausdorff metric in $\mathcal{F}_c(\mathbb{C})$ to be in the following form

$$d(\tilde{Z}, \tilde{W}) = d(\tilde{X} + i\tilde{Y}, \tilde{U} + i\tilde{V}) = \sup_{r\in(0,1]} d_H(X_r + iY_r, U_r + iV_r),$$

where $\tilde{Z} = \tilde{X} + i\tilde{Y}, \tilde{W} = \tilde{U} + i\tilde{V} \in \mathcal{F}_c(\mathbb{C})$.

Theorem 2.1. If $\tilde{Z} = \tilde{X} + i\tilde{Y}, \tilde{W} = \tilde{U} + i\tilde{V} \in \mathcal{F}_c(\mathbb{C})$, then we have

$$\max\left\{d(\tilde{X}, \tilde{U}), d(\tilde{Y}, \tilde{V})\right\} \leq d(\tilde{Z}, \tilde{W}) \leq d(\tilde{X}, \tilde{U}) + d(\tilde{Y}, \tilde{V}).$$

Proof: We have

$$d\left(\tilde{Z},\tilde{W}\right)=d\left(\tilde{X}+i\tilde{Y},\tilde{U}+i\tilde{V}\right)=\sup_{r\in(0,1]} d_H\left(X_r+iY_r,U_r+iV_r\right)$$

$$=\sup_{r\in(0,1]}\max\left\{\sup_{z_1\in X_r+iY_r}\inf_{z_2\in U_r+iV_r} d\left(z_1,z_2\right),\sup_{z_2\in U_r+iV_r}\inf_{z_1\in X_r+iY_r} d\left(z_1,z_2\right)\right\}$$

$$=\sup_{r\in(0,1]}\max\left\{d\left(X_r^-+iY_r^-,U_r^-+iV_r^-\right),d\left(X_r^-+iY_r^+,U_r^-+iV_r^+\right),d\left(X_r^++iY_r^-,U_r^++iV_r^-\right),\right.$$
$$\left.d\left(X_r^++iY_r^+,U_r^++iV_r^+\right),\left|X_r^--U_r^-\right|,\left|X_r^+-U_r^+\right|,\left|Y_r^--V_r^-\right|,\left|Y_r^+-V_r^+\right|\right\}.$$

Because

$$\max\left\{\sup_{r\in(0,1]}\max\left\{\left|X_r^--U_r^-\right|,\left|X_r^+-U_r^+\right|\right\},\sup_{r\in(0,1]}\max\left\{\left|Y_r^--V_r^-\right|,\left|Y_r^+-V_r^+\right|\right\}\right\}\le d\left(\tilde{Z},\tilde{W}\right),$$

we have

$$\max\left\{d\left(\tilde{X},\tilde{U}\right),d\left(\tilde{Y},\tilde{V}\right)\right\}\le d\left(\tilde{Z},\tilde{W}\right).$$

Because

$$d\left(\tilde{Z},\tilde{W}\right)\le\sup_{r\in(0,1]}\left\{\max\left\{\left|X_r^--U_r^-\right|,\left|X_r^+-U_r^+\right|\right\}+\max\left\{\left|Y_r^--V_r^-\right|,\left|Y_r^+-V_r^+\right|\right\}\right\}$$
$$\le\sup_{r\in(0,1]}\max\left\{\left|X_r^--U_r^-\right|,\left|X_r^+-U_r^+\right|\right\}+\sup_{r\in(0,1]}\max\left\{\left|Y_r^--V_r^-\right|,\left|Y_r^+-V_r^+\right|\right\}$$
$$=d_H\left(X,U\right)+d_H\left(Y,V\right),$$

we have

$$d\left(\tilde{Z},\tilde{W}\right)\le d\left(\tilde{X},\tilde{U}\right)+d\left(\tilde{Y},\tilde{V}\right).$$

The theorem is proven.

Definition 2.3 [38]**.** Let $(\Omega,\mathcal{A})$ be a measurable space, U be a universe of discourse, and $\left(\mathcal{F}\left(U\right),\mathcal{D}\right)$ be a fuzzy measurable space. $\tilde{F}$ is a fuzzy set valued mapping from Ω to $\mathcal{F}\left(U\right)$. If $\forall\tilde{D}\in\mathcal{D}$ and $\forall r\in[0,1]$, the following relation

$$F_r^{-1}\left(D_r\right)=\left\{\omega\in\Omega\left|F_r\left(\omega\right)\bigcap D_r\neq\Phi\right.\right\}\in\mathcal{A}$$

is valid, where $\mathcal{D}$ is a fuzzy $\sigma-$algebra of some fuzzy subsets of U, Φ is refereed as the empty set and $F_r\left(\omega\right)=\left\{u\in U\left|\tilde{F}\left(\omega\right)\left(u\right)\ge r\right.\right\}$, then we say $\tilde{F}$ is a measurable fuzzy set valued mapping from $(\Omega,\mathcal{A})$ to $\left(\mathcal{F}\left(U\right),\mathcal{D}\right)$.

Let $\mathbb{R}$ be the real line. $\mathfrak{B}$ is composed of all the Borel sets of $\mathbb{R}$. $I=[0,1]$. $\mathfrak{B}_I=\mathfrak{B}\bigcap I$ consists of all the Borel sets of I. We denotes

$$\tilde{\mathfrak{B}}=\left\{\tilde{B}\left|\tilde{B}\in\mathcal{F}\left(\mathbb{R}\right),B_r^{-1}\left(\mathfrak{B}_I\right)\subseteq\mathfrak{B},r\in I\right.\right\}.$$

Let $\mathbb{C}$ be the complex plane. $\mathfrak{B}_z$ consists of all the complex Borel sets of $\mathbb{C}$. We denotes

$$\tilde{\mathfrak{B}}_z=\left\{\tilde{C}=\tilde{A}+i\tilde{B}\left|\tilde{A},\tilde{B}\in\mathcal{F}\left(\mathbb{R}\right),A_r^{-1}\left(\mathfrak{B}_I\right),B_r^{-1}\left(\mathfrak{B}_I\right)\subseteq\mathfrak{B},r\in I\right.\right\}.$$

Definition 2.4 [38]. If $\tilde{F}$ is a measurable fuzzy set valued mapping from (Ω,A) to $\left(\mathcal{F}\left(\mathbb{R}\right),\tilde{\mathfrak{B}}\right)$, then we say $\tilde{F}$ is a measurable real fuzzy set valued function; If $\tilde{F}$ is a measurable fuzzy set valued mapping from $(\Omega,\mathcal{A})$ to $\left(\mathcal{F}\left(\mathbb{C}\right),\tilde{\mathfrak{B}}_z\right)$, then we say $\tilde{F}$ is a measurable complex fuzzy set valued function.

Definition 2.5 [38]**.** Let $(\Omega,\mathcal{A},P)$ be a probability measure space. A fuzzy set valued mapping $\tilde{\xi}:\Omega\to\mathcal{F}_c\left(\mathbb{R}\right)$ is called real fuzzy random variable if $\tilde{\xi}$ is a measurable real fuzzy set valued function, i.e., for $\forall r\in(0,1],\forall B\in\mathfrak{B}$,

$$\left\{\omega \in \Omega \middle| \xi_r(\omega) \bigcap B \neq \Phi\right\} \in \mathcal{A}.$$

Definition 2.6 [38]**.** Let $(\Omega, \mathcal{A}, P)$ is a probability measure space. A complex fuzzy set valued mapping

$$\tilde{\xi}: \Omega \to \mathcal{F}_c(\mathbb{C}) = \left\{\tilde{C} = \tilde{A} + i\tilde{B} \middle| \tilde{A}, \tilde{B} \in \mathcal{F}_c(\mathbb{R})\right\},$$

$$\omega \to \tilde{\xi}(\omega) = \tilde{\eta}(\omega) + i\tilde{\zeta}(\omega)$$

is called complex fuzzy random variable if $\tilde{\xi}$ is a measurable complex fuzzy set valued function defined on $(\Omega, \mathcal{A}, P)$. $\tilde{\eta}$ is a real part of $\tilde{\xi}$, $\tilde{\eta} = \mathrm{Re}(\tilde{\xi})$ and $\tilde{\zeta}$ is an imaginary part of $\tilde{\xi}$, that is $\tilde{\zeta} = \mathrm{Im}(\tilde{\xi})$.

Theorem 2.2 [38]. $\tilde{\xi} = \tilde{\eta} + i\tilde{\zeta}$ is a complex fuzzy random variable defined on $(\Omega, \mathcal{A}, P)$ if and only if $\tilde{\eta}$ and $\tilde{\zeta}$ are the real fuzzy random variables defined on $(\Omega, \mathcal{A}, P)$.

Definition 2.7 [16]**.** A real fuzzy random variable $\tilde{\xi}(\omega) = \left\{\left[\xi_r^-(\omega), \xi_r^+(\omega)\right] \middle| 0 \le r \le 1\right\}$ is called integrable if for each $r \in [0,1]$, $\xi_r^-(\omega)$ and $\xi_r^+(\omega)$ are integrable. In this case, the mathematical expectation of $\tilde{\xi}$ is defined in the following manner

$$E(\tilde{\xi}) = \int \tilde{\xi} dP = \left\{\left[\int \xi_r^- dP, \int \xi_r^+ dP\right] \middle| 0 \le r \le 1\right\}.$$

Definition 2.8 [38]. Let $\tilde{\xi}$ be a complex fuzzy random variable on $(\Omega, \mathcal{A}, P)$. We call

$$E(\tilde{\xi}) = \int \tilde{\xi} dP = \int \mathrm{Re}\,\tilde{\xi} dP + i\int \mathrm{Im}\,\tilde{\xi} dP = E(\mathrm{Re}\,\tilde{\xi}) + iE(\mathrm{Im}\,\tilde{\xi})$$

the expectation of $\tilde{\xi}$ if $E(\mathrm{Re}\,\tilde{\xi})$ and $E(\mathrm{Im}\,\tilde{\xi})$ both exist.

Theorem 2.3 [38]. If $\tilde{\xi}$ is a real (or complex) fuzzy random variable, then the following equalities

1. $\left\{E(\tilde{\xi})\right\}_r = E(\xi_r) = \left[E(\xi_r^-), E(\xi_r^+)\right]$ for $0 < r \le 1$;
2. $E(c\tilde{\xi}) = cE\tilde{\xi}$, whenever $c \in \mathbb{C}$;
3. $E(\tilde{\xi}_1 \pm \tilde{\xi}_2) = E\tilde{\xi}_1 \pm E\tilde{\xi}_2$

hold true.

Definition 2.9. Suppose that $\left\{\tilde{\xi}^{(t)}, t \in T\right\}$ is a family of real fuzzy random variables and T is any set of indexes.

1) If for any positive integer $n, (t_1, t_2, \cdots, t_n) \subseteq T$, the $\sigma-$algebra family $\left\{\sigma\left(\tilde{\xi}^{(t_s)}\right), s = 1, 2, \cdots, n\right\}$ is mutual independent, then we say that $\left\{\tilde{\xi}^{(t)}, t \in T\right\}$ is a family of mutual independent real fuzzy random variables. If T is a countable set, that is, $T = \{1, 2, \cdots\}$, then we say $\left\{\tilde{\xi}^{(k)}, k \in T\right\}$ is a sequence of independent real fuzzy random variables.

2) If for any positive integer $n, (t_1, t_2, \cdots, t_n) \subseteq T$ and $\forall r \in (0,1]$, the $\sigma-$ algebra family $\left\{\sigma\left(\left[\left(\tilde{\xi}^{(t_s)}\right)_r^-, \left(\tilde{\xi}^{(t_s)}\right)_r^+\right]\right), s = 1, 2, \cdots, n\right\}$ is mutual independent, then we say that $\left\{\tilde{\xi}^{(t)}, t \in T\right\}$ is a family of level-wise independent real fuzzy random variables. If T is a countable set, that is, $T = \{1, 2, \cdots\}$, then we say $\left\{\tilde{\xi}^{(k)}, k \in T\right\}$ is a sequence of level-wise independent real fuzzy random variables.

3) If $\forall r \in [0,1]$, $\left\{\left(\left(\tilde{\xi}^{(t)}\right)_r^-, \left(\tilde{\xi}^{(t)}\right)_r^+\right), t \in T\right\}$ is a family of identically distributed real random vectors, then we say that $\left\{\tilde{\xi}^{(t)}, t \in T\right\}$ is a family of level-wise identically distributed real fuzzy random variables.

Definition 2.10. Suppose that $\left\{\tilde{\xi}^{(t)}, t \in T\right\}$ is a family of complex fuzzy random variables, T is any set of indexes, and $\tilde{\xi}^{(t)} = \tilde{\eta}^{(t)} + i\tilde{\zeta}^{(t)}, t \in T$.

1) If $\left\{\overline{\xi}^{(t)} = \left(\tilde{\eta}^{(t)}, \tilde{\zeta}^{(t)}\right), t \in T\right\}$ is a family of independent (or level-wise independent) real fuzzy random vectors, then we say that $\left\{\xi^{(t)}, t \in T\right\}$ is a family of independent (or level-wise independent) complex fuzzy random variables.

2) If $\left\{\tilde{\eta}^{(t)}, t \in T\right\}$ is a family of level-wise identically distributed real fuzzy random variables, and $\left\{\zeta^{(t)}, t \in T\right\}$ is also a family of level-wise identically distributed real fuzzy random variables, then we say that $\left\{\xi^{(t)}, t \in T\right\}$ is a family of level-wise identically distributed complex fuzzy random variables.

Theorem 2.4 [38]**.** Let $\tilde{G}$ be a measurable complex fuzzy set valued function defined on $(\mathbb{C}, \mathfrak{B}_z)$, and $\tilde{\xi}$ be a complex fuzzy random variable defined on $(\Omega, \mathcal{A}, P)$. If $P\{A\} = 0$, where

$$A = \left\{\omega \in \Omega \middle| \forall r \in (0,1], \left\|G_r\left(\xi_r(\omega)\right)\right\| = \infty\right\},$$

then $\tilde{G}\left(\tilde{\xi}\right)$ is also a complex fuzzy random variable on $(\Omega, \mathcal{A}, P)$.

Theorem 2.5. Suppose that $\tilde{\xi}_j = \tilde{\eta}_j + i\tilde{\zeta}_j,\ j = 1, 2, \cdots, l$ is a sequence of level-wise independent and level-wise identically distributed complex fuzzy random variables. Let $\tilde{\xi}_j,\ j = 1, 2, \cdots, l$ be totally bounded, that is, there exist $A_1, B_1, A_2, B_2 \in \mathbb{R}$ such that

$$A_1 \le \left(\tilde{\eta}_j\right)_r \le B_1 \text{ and } A_2 \le \left(\tilde{\zeta}_j\right)_r \le B_2, \forall r \in (0,1],\ j = 1, 2, \cdots, l .$$

We also use the notation $A = \min\{A_1, A_2\}$ and $B = \max\{B_1, B_2\}$. The following inequality

$$P\left\{d\left(E\left(\frac{1}{l}\sum_{j=1}^{l}\tilde{\xi}_j\right), \frac{1}{l}\sum_{j=1}^{l}\tilde{\xi}_j\right) \ge \varepsilon\right\} < 4\exp\left\{-\frac{\varepsilon^2 l}{2(B-A)^2}\right\}$$

holds true.

Proof:

$$\begin{aligned}
&P\left\{d\left(E\left(\frac{1}{l}\sum_{j=1}^{l}\tilde{\xi}_j\right), \frac{1}{l}\sum_{j=1}^{l}\tilde{\xi}_j\right) \ge \varepsilon\right\} \\
&= P\left\{d\left(E\left[\frac{1}{l}\sum_{j=1}^{l}\tilde{\eta}_j\right] + iE\left[\frac{1}{l}\sum_{j=1}^{l}\tilde{\zeta}_j\right], \frac{1}{l}\sum_{j=1}^{l}\tilde{\eta}_j + i\frac{1}{l}\sum_{j=1}^{l}\tilde{\zeta}_j\right) \ge \varepsilon\right\} \\
&\le P\left\{\left\{d\left(E\left[\frac{1}{l}\sum_{j=1}^{l}\tilde{\eta}_j\right], \frac{1}{l}\sum_{j=1}^{l}\tilde{\eta}_j\right) + d\left(E\left[\frac{1}{l}\sum_{j=1}^{l}\tilde{\zeta}_j\right], \frac{1}{l}\sum_{j=1}^{l}\tilde{\zeta}_j\right)\right\} \ge \varepsilon\right\} \\
&\le P\left\{d\left(E\left[\frac{1}{l}\sum_{j=1}^{l}\tilde{\eta}_j\right], \frac{1}{l}\sum_{j=1}^{l}\tilde{\eta}_j\right) \ge \frac{\varepsilon}{2}\right\} + P\left\{d\left(E\left[\frac{1}{l}\sum_{j=1}^{l}\tilde{\zeta}_j\right], \frac{1}{l}\sum_{j=1}^{l}\tilde{\zeta}_j\right) \ge \frac{\varepsilon}{2}\right\} \\
&\le P\left\{\sup_{0<r\le 1} d_H\left(\left\{E\left[\frac{1}{l}\sum_{j=1}^{l}\tilde{\eta}_j\right]\right\}_r, \left\{\frac{1}{l}\sum_{j=1}^{l}\tilde{\eta}_j\right\}_r\right) \ge \frac{\varepsilon}{2}\right\} + P\left\{\sup_{0<r\le 1} d_H\left(\left\{E\left[\frac{1}{l}\sum_{j=1}^{l}\tilde{\zeta}_j\right]\right\}_r, \left\{\frac{1}{l}\sum_{j=1}^{l}\tilde{\zeta}_j\right\}_r\right) \ge \frac{\varepsilon}{2}\right\} \\
&\le P\left\{\sup_{0<r\le 1}\max\left(\left|\left\{E\left[\frac{1}{l}\sum_{j=1}^{l}\tilde{\eta}_j\right]\right\}_r^- - \left\{\frac{1}{l}\sum_{j=1}^{l}\tilde{\eta}_j\right\}_r^-\right|, \left|\left\{E\left[\frac{1}{l}\sum_{j=1}^{l}\tilde{\eta}_j\right]\right\}_r^+ - \left\{\frac{1}{l}\sum_{j=1}^{l}\tilde{\eta}_j\right\}_r^+\right|\right) \ge \frac{\varepsilon}{2}\right\} \\
&+ P\left\{\sup_{0<r\le 1}\max\left(\left|\left\{E\left[\frac{1}{l}\sum_{j=1}^{l}\tilde{\zeta}_j\right]\right\}_r^- - \left\{\frac{1}{l}\sum_{j=1}^{l}\tilde{\zeta}_j\right\}_r^-\right|, \left|\left\{E\left[\frac{1}{l}\sum_{j=1}^{l}\tilde{\zeta}_j\right]\right\}_r^+ - \left\{\frac{1}{l}\sum_{j=1}^{l}\tilde{\zeta}_j\right\}_r^+\right|\right) \ge \frac{\varepsilon}{2}\right\}
\end{aligned}$$

$$= P\left\{\sup_{0<r\le 1}\max\left(\left|E\left\{\frac{1}{l}\sum_{j=1}^{l}\left\{\tilde{\eta}_j\right\}_r^-\right\}-\frac{1}{l}\sum_{j=1}^{l}\left\{\tilde{\eta}_j\right\}_r^-\right|,\left|E\left\{\frac{1}{l}\sum_{j=1}^{l}\left\{\tilde{\eta}_j\right\}_r^+\right\}-\frac{1}{l}\sum_{j=1}^{l}\left\{\tilde{\eta}_j\right\}_r^+\right|\right)\ge\frac{\varepsilon}{2}\right\}$$

$$+P\left\{\sup_{0<r\le 1}\max\left(\left|E\left\{\frac{1}{l}\sum_{j=1}^{l}\left\{\tilde{\zeta}_j\right\}_r^-\right\}-\frac{1}{l}\sum_{j=1}^{l}\left\{\tilde{\zeta}_j\right\}_r^-\right|,\left|E\left\{\frac{1}{l}\sum_{j=1}^{l}\left\{\tilde{\zeta}_j\right\}_r^+\right\}-\frac{1}{l}\sum_{j=1}^{l}\left\{\tilde{\zeta}_j\right\}_r^+\right|\right)\ge\frac{\varepsilon}{2}\right\}$$

$$< 2\exp\left\{-\frac{\varepsilon^2 l}{2\left(B_1-A_1\right)^2}\right\}+2\exp\left\{-\frac{\varepsilon^2 l}{2\left(B_2-A_2\right)^2}\right\}< 4\exp\left\{-\frac{\varepsilon^2 l}{2\left(B-A\right)^2}\right\}.$$

3. Learning capability for the set of measurable complex fuzzy set valued functions

In this section, we will introduce concepts of capacity of the set of measurable complex fuzzy set valued functions. In these concepts, annealed entropy and growth function are non constructive, and VC dimension is constructive. Given these concepts, we will obtain expressions for the bounds of the rate of uniform convergence of learning process.

Definition 3.1. Let $\tilde{Z}_i = \tilde{X}_i + i\tilde{Y}_i,\ i = 1,2,\cdots,l$ be a sequence of level-wise independent and level-wise identically distributed complex fuzzy random variables, and let $\tilde{Q}\left(\tilde{Z},\alpha\right),\alpha\in\Lambda$ be a set of measurable complex fuzzy set valued functions. We call $\tilde{R}_{cf}\left(\alpha\right)=E\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]=\int\tilde{Q}\left(\tilde{Z},\alpha\right)dP$ the expected risk functional based on complex fuzzy random samples. It could be referred to as the complex fuzzy expected risk functional. We call the expression $\tilde{R}_{cfemp}\left(\alpha\right)=\frac{1}{l}\sum_{j=1}^{l}\tilde{Q}\left(\tilde{Z}_j,\alpha\right),\alpha\in\Lambda$ the complex fuzzy empirical risk functional.

Definition 3.2. Let $\tilde{Z}_j = \tilde{X}_j + i\tilde{Y}_j,\ j = 1,2,\cdots,l$ be a sequence of level-wise independent and level-wise identically distributed complex fuzzy random variables, and let $Q\left(\tilde{Z},\alpha\right),\alpha\in\Lambda$ be a set of measurable complex fuzzy set valued functions, Λ be any set of index. If there exists $\alpha_0\in\Lambda$ such that the relation

$$\left\|\tilde{R}_{cf}\left(\alpha_0\right)\right\| = \inf_{\alpha\in\Lambda}\left\|\tilde{R}_{cf}\left(\alpha\right)\right\|$$

is valid, then we say that $\tilde{R}_{cf}\left(\alpha_0\right)$ is the greatest lower bound of $\tilde{R}_{cf}\left(\alpha\right)$, denoted by

$$\tilde{R}_{cf}\left(\alpha_0\right) = \inf_{\alpha\in\Lambda}\tilde{R}_{cf}\left(\alpha\right).$$

Similarly, if there exists $\alpha_l\in\Lambda$ such that the relation

$$\left\|\tilde{R}_{cfemp}\left(\alpha_l\right)\right\| = \inf_{\alpha\in\Lambda}\left\|\tilde{R}_{cfemp}\left(\alpha\right)\right\|$$

is valid, then we say that $\tilde{R}_{cfemp}\left(\alpha_l\right)$ is the greatest lower bound of $\tilde{R}_{cfemp}\left(\alpha\right)$, denoted by

$$\tilde{R}_{cfemp}\left(\alpha_l\right) = \inf_{\alpha\in\Lambda}\tilde{R}_{cfemp}\left(\alpha\right).$$

The principle of empirical risk minimization can be described in the following manner. Let us, instead of minimizing the complex fuzzy expected risk functional $\tilde{R}_{cf}\left(\alpha\right)$, minimize the complex fuzzy empirical risk functional $\tilde{R}_{cfemp}\left(\alpha\right)$. Consider that the minimum of the complex fuzzy expected risk functional is attained at $\tilde{Q}\left(\tilde{Z},\alpha_0\right)$ and suppose that the minimum of the complex fuzzy empirical risk functional is attained at $\tilde{Q}\left(\tilde{Z},\alpha_l\right)$. We view the function $\tilde{Q}\left(\tilde{Z},\alpha_l\right)$ as an approximation of the original function $\tilde{Q}\left(\tilde{Z},\alpha_0\right)$. The principle of solving the risk minimization problem is called the complex fuzzy empirical risk minimization principle, the CFERM principle, to be brief.

Definition 3.3. Let $\tilde{Q}\left(\tilde{X},\alpha\right),\alpha\in\Lambda$ be a set of measurable real fuzzy set valued functions, and $\tilde{X}$ be a real fuzzy random variable. We call the set of indicator functions $\theta\left(\left[\tilde{Q}\left(\tilde{X},\alpha\right)\right]_r^- - \beta\right),\ \alpha\in\Lambda$,

$$\beta \in \boldsymbol{\mathcal{B}}_{\tilde{X}}^{-} = \left(\inf_{\tilde{X} \in \mathcal{F}_c(\mathbb{R}), \alpha \in \Lambda} \left[\tilde{Q}\left(\tilde{X}, \alpha\right) \right]_r^{-}, \sup_{\tilde{X} \in \mathcal{F}_c(\mathbb{R}), \alpha \in \Lambda} \left[\tilde{Q}\left(\tilde{X}, \alpha\right) \right]_r^{-} \right),\ r \in [0,1]$$

where

$$\theta(u) = \begin{cases} 1, u \geq 0 \\ 0, u < 0 \end{cases},$$

the left complete set of indicators for a set of measurable real fuzzy set valued functions $Q\left(\tilde{X}, \alpha\right), \alpha \in \Lambda$.

Definition 3.4. Let $Q\left(\tilde{X}, \alpha\right), \tilde{X} \in \mathcal{F}_c(\mathbb{R}), \alpha \in \Lambda$ be a set of measurable real fuzzy set valued functions. Let $N_{-}^{\Lambda,\beta}\left(\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_l\right)$ be the number of different separations of l vectors $\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_l$ by a left complete set of indicators:

$$\theta\left(\left[\tilde{Q}\left(\tilde{X}, \alpha\right) \right]_r^{-} - \beta \right),$$

$$\alpha \in \Lambda,\ \beta \in \boldsymbol{\mathcal{B}}_{\tilde{X}}^{-} = \left(\inf_{\tilde{X} \in \mathcal{F}_c(\mathbb{R}), \alpha \in \Lambda} \left[\tilde{Q}\left(\tilde{X}, \alpha\right) \right]_r^{-}, \sup_{\tilde{X} \in \mathcal{F}_c(\mathbb{R}), \alpha \in \Lambda} \left[\tilde{Q}\left(\tilde{X}, \alpha\right) \right]_r^{-} \right),\ r \in [0,1].$$

Let the function

$$H_{-}^{\Lambda,\beta}\left(\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_l\right) = \ln N_{-}^{\Lambda,\beta}\left(\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_l\right)$$

be measurable with respect to probability P on $\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_l$.

We call the quantity

$$\left(H_{ann}^{\Lambda, \boldsymbol{\mathcal{B}}_{\tilde{X}}^{-}} \right)(l) = \ln E N_{-}^{\Lambda,\beta}\left(\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_l\right)$$

the left annealed entropy of the set indicators of measurable real fuzzy set valued functions.

Definition 3.5. We call the quantity

$$G^{\Lambda, \boldsymbol{\mathcal{B}}_{\tilde{X}}^{-}}(l) = \ln\left[\max_{\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_l} N_{-}^{\Lambda,\beta}\left(\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_l\right) \right]$$

the left growth function of a set of measurable real fuzzy set valued functions $\tilde{Q}\left(\tilde{X}, \alpha\right), \tilde{X} \in \mathcal{F}_c(\mathbb{R}), \alpha \in \Lambda$.

Definition 3.6. We call the maximal number h_{-} of vector $\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_l$ that can be shattered by the left complete set of indicators

$$\theta\left(\left[\tilde{Q}\left(\tilde{X}, \alpha\right) \right]_r^{-} - \beta \right),$$

$$\alpha \in \Lambda,\ \beta \in \boldsymbol{\mathcal{B}}_{\tilde{X}}^{-} = \left(\inf_{\tilde{X} \in \mathcal{F}_c(\mathbb{R}), \alpha \in \Lambda} \left[\tilde{Q}\left(\tilde{X}, \alpha\right) \right]_r^{-}, \sup_{\tilde{X} \in \mathcal{F}_c(\mathbb{R}), \alpha \in \Lambda} \left[\tilde{Q}\left(\tilde{X}, \alpha\right) \right]_r^{-} \right),\ r \in [0,1],$$

the left VC dimension of the set of measurable real fuzzy set valued functions $\tilde{Q}\left(\tilde{X}, \alpha\right), \alpha \in \Lambda$.

Similarly, we have the following definitions:

Definition 3.7. Let $\tilde{Q}\left(\tilde{X}, \alpha\right), \alpha \in \Lambda$ be a set of measurable real fuzzy set valued functions, and $\tilde{X}$ be a real fuzzy random variable. We call the set of indicator functions

$$\theta\left(\left[\tilde{Q}\left(\tilde{X}, \alpha\right) \right]_r^{+} - \beta \right),$$

$$\alpha \in \Lambda,\ \beta \in \boldsymbol{\mathcal{B}}_{\tilde{X}}^{+} = \left(\inf_{\tilde{X} \in \mathcal{F}_c(\mathbb{R}), \alpha \in \Lambda} \left[\tilde{Q}\left(\tilde{X}, \alpha\right) \right]_r^{+}, \sup_{\tilde{X} \in \mathcal{F}_c(\mathbb{R}), \alpha \in \Lambda} \left[\tilde{Q}\left(\tilde{X}, \alpha\right) \right]_r^{+} \right),\ r \in [0,1],$$

the right complete set of indicators for a set of measurable real fuzzy set valued functions $\tilde{Q}\left(\tilde{X}, \alpha\right), \alpha \in \Lambda$.

Definition 3.8. Let $\tilde{Q}\left(\tilde{X}, \alpha\right), \tilde{X} \in, \alpha \in \Lambda$ be a set of measurable real fuzzy set valued functions. Let $N_{+}^{\Lambda,\beta}\left(\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_l\right)$ be the number of different separations of l vectors $\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_l$ by a right complete set of indicators:

$$\theta\left(\left[\tilde{Q}\left(\tilde{X},\alpha\right)\right]_r^+ - \beta\right),$$

$$\alpha \in \Lambda\,,\ \beta \in \mathcal{B}_{\tilde{X}}^+ = \left(\inf_{\tilde{X}\in\mathcal{F}_c(\mathbb{R}),\alpha\in\Lambda}\left[\tilde{Q}\left(\tilde{X},\alpha\right)\right]_r^+,\ \sup_{\tilde{X}\in\mathcal{F}_c(\mathbb{R}),\alpha\in\Lambda}\left[\tilde{Q}\left(\tilde{X},\alpha\right)\right]_r^+\right),\ r\in[0,1].$$

Let the function

$$H_+^{\Lambda,\beta}\left(\tilde{X}_1,\tilde{X}_2,\cdots,\tilde{X}_l\right) = \ln N_+^{\Lambda,\beta}\left(\tilde{X}_1,\tilde{X}_2,\cdots,\tilde{X}_l\right)$$

be measurable with respect to probability P on $\tilde{X}_1,\tilde{X}_2,\cdots,\tilde{X}_l$.

We call the quantity

$$\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{X}}^+}\right)(l) = \ln EN_+^{\Lambda,\beta}\left(\tilde{X}_1,\tilde{X}_2,\cdots,\tilde{X}_l\right)$$

the right annealed entropy of the set indicators of measurable real fuzzy set valued functions.

Definition 3.9. We call the quantity

$$G^{\Lambda,\mathcal{B}_{\tilde{X}}^+}(l) = \ln\left[\max_{\tilde{X}_1,\tilde{X}_2,\cdots,\tilde{X}_l} N_+^{\Lambda,\beta}\left(\tilde{X}_1,\tilde{X}_2,\cdots,\tilde{X}_l\right)\right]$$

the right growth function of a set of measurable real fuzzy set valued functions $\tilde{Q}\left(\tilde{X},\alpha\right),\tilde{X}\in\mathcal{F}_c(\mathbb{R}),\alpha\in\Lambda$.

Definition 3.10. We call the maximal number h_+ of vector $\tilde{X}_1,\tilde{X}_2,\cdots,\tilde{X}_l$ that can be shattered by the complete set of indicators

$$\theta\left(\left[\tilde{Q}\left(\tilde{X},\alpha\right)\right]_r^+ - \beta\right),$$

$$\alpha \in \Lambda\,,\ \beta \in \mathcal{B}_{\tilde{X}}^+ = \left(\inf_{\tilde{X}\in\mathcal{F}_c(\mathbb{R}),\alpha\in\Lambda}\left[\tilde{Q}\left(\tilde{X},\alpha\right)\right]_r^+,\ \sup_{\tilde{X}\in\mathcal{F}_c(\mathbb{R}),\alpha\in\Lambda}\left[\tilde{Q}\left(\tilde{X},\alpha\right)\right]_r^+\right),\ r\in[0,1],$$

the right VC dimension of the set of measurable real fuzzy set valued functions $\tilde{Q}\left(\tilde{X},\alpha\right),\alpha\in\Lambda$.

Definition 3.11. We call respectively

$$\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{X}}}\right)(l) = \max\left\{\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{X}}^-}\right)(l),\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{X}}^+}\right)(l)\right\},$$

$$G^{\Lambda,\mathcal{B}_{\tilde{X}}}(l) = \max\left\{G^{\Lambda,\mathcal{B}_{\tilde{X}}^-}(l),G^{\Lambda,\mathcal{B}_{\tilde{X}}^+}(l)\right\},$$

$$h_{\tilde{X}} = \max\{h_-,h_+\}$$

the annealed entropy, the growth function and the VC dimension of a set of measurable real fuzzy set valued functions $\tilde{Q}\left(\tilde{X},\alpha\right),\alpha\in\Lambda$.

Definition 3.12. Let $\tilde{Z} = \tilde{X} + i\tilde{Y}$ be a complex fuzzy random sample, and let

$$\tilde{Q}\left(\tilde{Z},\alpha\right) = \mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right] + i\,\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right],\alpha\in\Lambda$$

be a set of measurable complex fuzzy set valued functions. Suppose that the annealed entropy, the growth function and VC dimension of the set of measurable complex fuzzy set valued functions $Q\left(\tilde{Z},\alpha\right),\alpha\in\Lambda$ are respectively $\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{Z}}}\right)(l)$ 、 $G^{\Lambda,\mathcal{B}_{\tilde{Z}}}(l)$ and $h_{\tilde{Z}}$, the annealed entropy, the growth function and VC dimension of the real parts $\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right],\alpha\in\Lambda$ of $Q\left(\tilde{Z},\alpha\right),\alpha\in\Lambda$ are respectively $\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{X}}}\right)(l)$ 、 $G^{\Lambda,\mathcal{B}_{\tilde{X}}}(l)$ and $h_{\tilde{X}}$, and the annealed entropy, the growth function and VC dimension of the imaginary parts $\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right],\alpha\in\Lambda$ of $Q\left(\tilde{Z},\alpha\right),\alpha\in\Lambda$ are respectively $\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{Y}}}\right)(l)$ 、 $G^{\Lambda,\mathcal{B}_{\tilde{Y}}}(l)$ and $h_{\tilde{Y}}$. We define that

$$\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{Z}}}\right)(l) = \max\left\{\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{X}}}\right)(l),\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{Y}}}\right)(l)\right\},$$

$$G^{\Lambda,\mathcal{B}_{\tilde{Z}}}(l)=\max\left\{G^{\Lambda,\mathcal{B}_{\tilde{X}}}(l),G^{\Lambda,\mathcal{B}_{\tilde{Y}}}(l)\right\},$$

$$h_{\tilde{Z}}=\max\left\{h_{\tilde{X}},h_{\tilde{Y}}\right\}.$$

Theorem 3.1. The following inequalities

$$\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{X}}}\right)(l)\le G^{\Lambda,\mathcal{B}_{\tilde{X}}}(l)\begin{cases}=l\ln 2 & \text{if } \ l\le h_{\tilde{X}},\\ \le h_{\tilde{X}}\left(1+\ln\dfrac{l}{h_{\tilde{X}}}\right) & \text{if } \ l>h_{\tilde{X}},\end{cases},$$

$$\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{Y}}}\right)(l)\le G^{\Lambda,\mathcal{B}_{\tilde{Y}}}(l)\begin{cases}=l\ln 2 & \text{if } \ l\le h_{\tilde{Y}},\\ \le h_{\tilde{Y}}\left(1+\ln\dfrac{l}{h_{\tilde{Y}}}\right) & \text{if } \ l>h_{\tilde{Y}},\end{cases},$$

$$\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{Z}}}\right)(l)\le G^{\Lambda,\mathcal{B}_{\tilde{Z}}}(l)\begin{cases}=l\ln 2 & \text{if } \ l\le h_{\tilde{Z}},\\ \le h_{\tilde{Z}}\left(1+\ln\dfrac{l}{h_{\tilde{Z}}}\right) & \text{if } \ l>h_{\tilde{Z}},\end{cases},$$

hold true.

The proof goes the same way as in the proof of Theorem 4.3 of Ref. [28] with the aid of the Definitions 3.3-3.12, and will be omitted.

Remark 3.1. In classical statistical learning theory, the concepts and some important properties (see Ref. [27, 28]) of the annealed entropy, the growth function and VC dimension of real measurable risk functions $Q(z,\alpha),\alpha\in\Lambda$ are the special cases of this study.

4. Determination of bounds on the rate of uniform convergence of learning process

In the SLT, the important conclusions about the relation between the empirical risk and practical risk are expressed in the form of the bounds of generalization. They become essential when analyzing the capacity of learning machines and developing new learning algorithms. The bounds on the rates of uniform convergence of learning process are important components of the bounds of generalization. In this section, we discuss the bounds on the rate of uniform convergence of learning process based on complex fuzzy random samples. To achieve this goal, we need employ the concepts proposed in Section 3. We will obtain expressions for the bounds of the rate of uniform convergence of learning process through the following two cases.

In this section, we use the notation $\delta(l)=4\dfrac{\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{Z}}}\right)(2l)-\ln\left(\dfrac{\eta}{8}\right)}{l}$.

4.1. $\left\{\tilde{Q}\left(\tilde{Z},\alpha\right),\alpha\in\Lambda\right\}$ is a set of totally bounded measurable complex fuzzy set valued functions.

Theorem 4.1. Suppose that $\tilde{Q}\left(\tilde{Z},\alpha\right),\alpha\in\Lambda$ is a set of totally bounded measurable complex fuzzy set valued functions, that is, there exist $A_1,B_1,A_2,B_2\in\mathbb{R}$ such that

$$A_1\le\left\{\operatorname{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right\}_r\le B_1 \text{ and } A_2\le\left\{\operatorname{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right\}_r\le B_2,\ \forall r\in(0,1],\forall\alpha\in\Lambda.$$

We also use the notation $A=\min\{A_1,A_2\}$ and $B=\max\{B_1,B_2\}$. The following inequality

$$P\left\{\sup_{\alpha\in\Lambda}d\left(\tilde{R}_{cf}(\alpha),\tilde{R}_{cfemp}(\alpha)\right)>\varepsilon\right\}\le 8\exp\left\{\left(\frac{\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{Z}}}\right)(2l)}{l}-\frac{\varepsilon^2}{4(B-A)^2}\right)l\right\} \tag{4.1}$$

holds true.

Proof:

$$P\left\{\sup_{\alpha\in\Lambda} d\left(\tilde{R}_{cf}(\alpha),\tilde{R}_{cfemp}(\alpha)\right)>\varepsilon\right\}=P\left\{\sup_{\alpha\in\Lambda} d\left(E\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right],\frac{1}{l}\sum_{j=1}^{l}\tilde{Q}\left(\tilde{Z}_j,\alpha\right)\right)>\varepsilon\right\}$$

$$\le P\left\{\sup_{\alpha\in\Lambda} d\left(E\left[\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right],\frac{1}{l}\sum_{j=1}^{l}\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z}_j,\alpha\right)\right]\right)\ge\frac{\varepsilon}{2}\right\}+P\left\{\sup_{\alpha\in\Lambda} d\left(E\left[\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right],\frac{1}{l}\sum_{j=1}^{l}\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z}_j,\alpha\right)\right]\right)\ge\frac{\varepsilon}{2}\right\}$$

$$\le P\left\{\sup_{\alpha\in\Lambda}\sup_{0<r\le 1} d_H\left(\left\{E\left[\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right]\right\}_r,\left\{\frac{1}{l}\sum_{j=1}^{l}\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z}_j,\alpha\right)\right]\right\}_r\right)\ge\frac{\varepsilon}{2}\right\}$$

$$+P\left\{\sup_{\alpha\in\Lambda}\sup_{0<r\le 1} d_H\left(\left\{E\left[\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right]\right\}_r,\left\{\frac{1}{l}\sum_{j=1}^{l}\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z}_j,\alpha\right)\right]\right\}_r\right)\ge\frac{\varepsilon}{2}\right\}$$

$$\le P\left\{\max\left(\sup_{\alpha\in\Lambda}\sup_{0<r\le 1}\left|E\left\{\left\{\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right\}_r^-\right\}-\frac{1}{l}\sum_{j=1}^{l}\left\{\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z}_j,\alpha\right)\right]\right\}_r^-\right|,\sup_{\alpha\in\Lambda}\sup_{0<r\le 1}\left|E\left\{\left\{\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right\}_r^+\right\}-\frac{1}{l}\sum_{j=1}^{l}\left\{\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z}_j,\alpha\right)\right]\right\}_r^+\right|\right)\ge\frac{\varepsilon}{2}\right\}$$

$$+P\left\{\max\left(\sup_{\alpha\in\Lambda}\sup_{0<r\le 1}\left|E\left\{\left\{\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right\}_r^-\right\}-\frac{1}{l}\sum_{j=1}^{l}\left\{\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z}_j,\alpha\right)\right]\right\}_r^-\right|,\sup_{\alpha\in\Lambda}\sup_{0<r\le 1}\left|E\left\{\left\{\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right\}_r^+\right\}-\frac{1}{l}\sum_{j=1}^{l}\left\{\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z}_j,\alpha\right)\right]\right\}_r^+\right|\right)\ge\frac{\varepsilon}{2}\right\}$$

$$\le 4\exp\left\{\left(\frac{\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{X}}}\right)(2l)}{l}-\frac{\varepsilon^2}{4\left(B_1-A_1\right)^2}\right)l\right\}+4\exp\left\{\left(\frac{\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{Y}}}\right)(2l)}{l}-\frac{\varepsilon^2}{4\left(B_2-A_2\right)^2}\right)l\right\}$$

$$\le 8\exp\left\{\left(\frac{\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{Z}}}\right)(2l)}{l}-\frac{\varepsilon^2}{4\left(B-A\right)^2}\right)l\right\}.$$

Theorem 4.2. For all functions in a set of measurable complex fuzzy set valued functions $\tilde{Q}\left(\tilde{Z},\alpha\right),\alpha\in\Lambda$ which satisfy the conditions for theorem 4.1, we have

(1) the following inequality

$$d\left(\tilde{R}_{cf}(\alpha),\tilde{R}_{cfemp}(\alpha)\right)\le\left(B-A\right)\sqrt{\delta(l)} \qquad (4.2)$$

holds true with probability $1-\eta$.

(2) the inequality

$$\Delta(\alpha_l)=\left\|\tilde{R}_{cf}(\alpha_l)\right\|-\left\|\tilde{R}_{cf}(\alpha_0)\right\|\le\left(B-A\right)\sqrt{\delta(l)+}\left(B-A\right)\sqrt{-\frac{2\ln\left(\frac{\eta}{4}\right)}{l}} \qquad (4.3)$$

holds with probability at least $1-2\eta$.

Proof. Let us rewrite the inequality (4.1) in a certain equivalent form. To do this we introduce a positive value $0<\eta\le 1$ and the equality

$$8\exp\left\{\left(\frac{\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{Z}}}\right)(2l)}{l}-\frac{\varepsilon^2}{4\left(B-A\right)^2}\right)l\right\}=\eta,$$

which we solve with respect to ε. We obtain

$$\varepsilon=\left(B-A\right)\sqrt{\delta(l)}.$$

Now the assertion comes in the following equivalent form:

With probability $1-\eta$ simultaneously for all functions in the set $\tilde{Q}\left(\tilde{Z},\alpha\right),\alpha\in\Lambda$, the inequality

$$d\left(\tilde{R}_{cf}(\alpha),\tilde{R}_{cfemp}(\alpha)\right)\le\left(B-A\right)\sqrt{\delta(l)}$$

is valid.

Let $\tilde{Q}\left(\tilde{Z},\alpha_0\right)$ be a function from the set of functions that minimizes the complex fuzzy expected risk functional $\tilde{R}_{cf}(\alpha)$ and let $\tilde{Q}\left(\tilde{Z},\alpha_l\right)$ be a function from this set that minimizes the complex fuzzy empirical risk functional $\tilde{R}_{cfemp}(\alpha)$. Since the inequality is true for all functions in the set, it is true as well for the function $\tilde{Q}\left(\tilde{Z},\alpha_l\right)$. Thus with probability at least $1-\eta$ the following inequalities

$$\left\|\tilde{R}_{cf}(\alpha_l)\right\| \le \left\|\tilde{R}_{cfemp}(\alpha_l)\right\| + (B-A)\sqrt{\delta(l)} \tag{4.4}$$

is valid.

For the function $\tilde{Q}\left(\tilde{Z},\alpha_0\right)$ which minimizes $\tilde{R}_{cf}(\alpha)$, according to Theorem 2.5, the following relationship

$$\left\|\tilde{R}_{cfemp}(\alpha_0)\right\| \le \left\|\tilde{R}_{cf}(\alpha_0)\right\| + (B-A)\sqrt{-\frac{2\ln\left(\frac{\eta}{4}\right)}{l}} \tag{4.5}$$

holds true with probability at least $1-\eta$.

According to (4.4) and (4.5), we conclude that with probability at least $1-2\eta$ the inequalities

$$\begin{aligned}\Delta(\alpha_l) &= \left\|\tilde{R}_{cf}(\alpha_l)\right\| - \left\|\tilde{R}_{cf}(\alpha_0)\right\| \\ &= \left\|\tilde{R}_{cf}(\alpha_l)\right\| - \left\|\tilde{R}_{cfemp}(\alpha_l)\right\| + \left\|\tilde{R}_{cfemp}(\alpha_l)\right\| - \left\|\tilde{R}_{cfemp}(\alpha_0)\right\| + \left\|\tilde{R}_{cfemp}(\alpha_0)\right\| - \left\|\tilde{R}_{cf}(\alpha_0)\right\| \\ &\le (B-A)\sqrt{\delta(l)} + (B-A)\sqrt{-\frac{2\ln\left(\frac{\eta}{4}\right)}{l}}\end{aligned}$$

are satisfied with probability at least $1-2\eta$.

4.2. $\left\{\tilde{Q}\left(\tilde{Z},\alpha\right),\alpha\in\Lambda\right\}$ is a set of totally bounded nonnegative measurable complex fuzzy set valued functions.

Theorem 4.3. Suppose that $\tilde{Q}\left(\tilde{Z},\alpha\right),\alpha\in\Lambda$ is a set of totally bounded nonnegative measurable complex fuzzy set valued functions, that is, there exist $B_1, B_2 \in \mathbb{R}$ such that

$$0 \le \left\{\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right\}_r \le B_1 \text{ and } 0 \le \left\{\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right\}_r \le B_2,\ \forall r\in(0,1],\forall\alpha\in\Lambda.$$

We also use the notation $B=\max\left\{B_1,B_2\right\}$. The following inequality

$$P\left\{\sup_{\alpha\in\Lambda}\frac{d\left(\tilde{R}_{cf}(\alpha),\tilde{R}_{cfemp}(\alpha)\right)}{\sqrt{\left\|\tilde{R}_{cf}(\alpha)\right\|}} > \varepsilon\right\} \le 8\exp\left\{\left(\frac{\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{Z}}}\right)(2l)}{l} - \frac{\varepsilon^2}{16B}\right)l\right\} \tag{4.6}$$

holds true.

Proof:

$$P\left\{\sup_{\alpha\in\Lambda}\frac{d\left(\tilde{R}_{cf}(\alpha),\tilde{R}_{cfemp}(\alpha)\right)}{\sqrt{\left\|\tilde{R}_{cf}(\alpha)\right\|}} > \varepsilon\right\}$$

$$\le P\left\{\sup_{\alpha\in\Lambda}\frac{d\left(E\left[\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right],\frac{1}{l}\sum_{j=1}^{l}\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z}_j,\alpha\right)\right]\right)}{\sqrt{\left\|\tilde{R}_{cf}(\alpha)\right\|}} > \frac{\varepsilon}{2}\right\} + P\left\{\sup_{\alpha\in\Lambda}\frac{d\left(E\left[\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right],\frac{1}{l}\sum_{j=1}^{l}\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z}_j,\alpha\right)\right]\right)}{\sqrt{\left\|\tilde{R}_{cf}(\alpha)\right\|}} > \frac{\varepsilon}{2}\right\}$$

$$\le P\left\{\max\left(\sup_{\alpha\in\Lambda}\sup_{0<r\le1}\frac{\left|E\left\{\left\{\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right\}_r^-\right\} - \frac{1}{l}\sum_{j=1}^{l}\left\{\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z}_j,\alpha\right)\right]\right\}_r^-\right|}{\sqrt{E\left\{\left\{\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right\}_r^-\right\}}}, \sup_{\alpha\in\Lambda}\sup_{0<r\le1}\frac{\left|E\left\{\left\{\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right\}_r^+\right\} - \frac{1}{l}\sum_{j=1}^{l}\left\{\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z}_j,\alpha\right)\right]\right\}_r^+\right|}{\sqrt{E\left\{\left\{\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right\}_r^+\right\}}}\right) > \frac{\varepsilon}{2}\right\}$$

$$+P\left\{\max\left(\sup_{\alpha\in\Lambda}\sup_{0<r\le 1}\frac{\left|E\left\{\left\{\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right\}_r^-\right\}-\frac{1}{l}\sum_{j=1}^{l}\left\{\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z}_j,\alpha\right)\right]\right\}_r^-\right|}{\sqrt{E\left\{\left\{\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right\}_r^-\right\}}},\sup_{\alpha\in\Lambda}\sup_{0<r\le 1}\frac{\left|E\left\{\left\{\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right\}_r^+\right\}-\frac{1}{l}\sum_{j=1}^{l}\left\{\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z}_j,\alpha\right)\right]\right\}_r^+\right|}{\sqrt{E\left\{\left\{\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha\right)\right]\right\}_r^+\right\}}}\right)>\frac{\varepsilon}{2}\right\}$$

$$<4\exp\left\{\left(\frac{\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{X}}}\right)(2l)}{l}-\frac{\varepsilon^2}{16B_1}\right)l\right\}+4\exp\left\{\left(\frac{\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{Y}}}\right)(2l)}{l}-\frac{\varepsilon^2}{16B_2}\right)l\right\}\le 8\exp\left\{\left(\frac{\left(H_{ann}^{\Lambda,\mathcal{B}_{\tilde{Z}}}\right)(2l)}{l}-\frac{\varepsilon^2}{16B}\right)l\right\}.$$

Theorem 4.4. For all functions in a set of measurable complex fuzzy set valued functions $\tilde{Q}\left(\tilde{Z},\alpha\right),\alpha\in\Lambda$ which satisfy the conditions for theorem 4.3, we have

(1) the following inequality

$$\left\|\tilde{R}_{cf}\left(\alpha\right)\right\|\le\left\|\tilde{R}_{cfemp}\left(\alpha\right)\right\|+2B\delta\left(l\right)\left[1+\sqrt{1+\frac{\left\|\tilde{R}_{cfemp}\left(\alpha\right)\right\|}{B\delta\left(l\right)}}\right] \tag{4.7}$$

holds true with probability $1-\eta$.

(2) the inequality

$$\Delta\left(\alpha_l\right)=\left\|\tilde{R}_{cf}\left(\alpha_l\right)\right\|-\left\|\tilde{R}_{cf}\left(\alpha_0\right)\right\|\le 2B\delta\left(l\right)\left[1+\sqrt{1+\frac{\left\|\tilde{R}_{cfemp}\left(\alpha\right)\right\|}{B\delta\left(l\right)}}\right]+B\sqrt{-\frac{2\ln\left(\frac{\eta}{4}\right)}{l}} \tag{4.8}$$

holds with probability at least $1-2\eta$.

Proof: We can prove this theorem in the same way as done in Theorem 4.2, and will be omitted.

Remark 4.1. The bounds described by the inequalities (4.2) and (4.7) depend on the probability distribution P. One can derive both distribution-free non constructive bounds and distribution-free constructive bounds. To obtain these bounds, it is sufficient in the inequalities (4.2) and (4.7) to use the expression

$$\delta\left(l\right)=4\frac{G^{\Lambda,\mathcal{B}_{\tilde{Z}}}\left(l\right)\left(2l\right)-\ln\left(\frac{\eta}{8}\right)}{l}$$

(this expression provides distribution-free non constructive bounds), or to use the expression

$$\delta\left(l\right)=4\frac{h_{\tilde{Z}}\left(\ln\frac{2l}{h_{\tilde{Z}}}+1\right)-\ln\left(\frac{\eta}{8}\right)}{l}$$

(this expression provides distribution-free constructive bounds).

5. Structural risk minimization principle based on complex fuzzy random samples

According to Theorem 4.2 (1), we obtained that with probability at least $1-\eta$ simultaneously for all functions in a set of measurable complex fuzzy set valued functions $\tilde{Q}\left(\tilde{Z},\alpha\right),\alpha\in\Lambda$, with finite VC dimension the inequality

$$\left\|\tilde{R}_{cf}\left(\alpha\right)\right\|\le\left\|\tilde{R}_{cfemp}\left(\alpha\right)\right\|+\left(B-A\right)\sqrt{4\frac{h_{\tilde{Z}}\left(\ln\frac{2l}{h_{\tilde{Z}}}+1\right)-\ln\left(\frac{\eta}{8}\right)}{l}} \tag{5.1}$$

holds true.

According to Theorem 4.4 (1), We obtained also that with probability at least $1-\eta$ simultaneously for all functions in a set of measurable complex fuzzy set valued functions $\tilde{Q}\left(\tilde{Z},\alpha\right),\alpha\in\Lambda$, the inequality

$$\left\|\tilde{R}_{cf}(\alpha)\right\| \le \left\|\tilde{R}_{cfemp}(\alpha)\right\| + 2B\delta(l)\left[1+\sqrt{1+\frac{\left\|\tilde{R}_{cfemp}(\alpha)\right\|}{4B\frac{h_{\tilde{Z}}\left(\ln\frac{2l}{h_{\tilde{Z}}}+1\right)-\ln\left(\frac{\eta}{8}\right)}{l}}}\right] \tag{5.2}$$

holds true.

According to inequalities (5.1) and (5.2) the upper bound on the risk decreases with decreasing the value of empirical risk. This is the reason why the principle of complex fuzzy empirical risk minimization often gives good results for large sample size. However, if ratio of the number of the training patterns to the VC dimension of the set of functions of the learning machines is small, a small value of the complex fuzzy empirical risk $\tilde{R}_{cfemp}(\alpha)$ does not guarantee a small value of the complex fuzzy actual risk. In this case, to minimize the complex fuzzy actual risk $\tilde{R}_{cf}(\alpha)$ one has to minimize the right-hand side of inequality (5.1) (or (5.2)) simultaneously over both terms. Note that the first term in inequality (5.1) depends on a specific function of the set of functions, while for a fixed number of observations the second term depends mainly on the VC dimension of the whole set of measurable complex fuzzy set valued functions. Therefore to minimize the right-hand side of the bound of risk, (5.1) (or (5.2)), simultaneously over both terms, one has to make the VC dimension a controlling variable.

To do this we consider the following scheme.

Let us impose the structure $\mathbb{S}$ on the set S of measurable complex fuzzy set valued functions $\tilde{Q}(\tilde{Z},\alpha),\alpha\in\Lambda$. Consider the set of nested subsets of functions

$$S_1 \subset S_2 \subset \cdots \subset S_n, \cdots, \tag{5.3}$$

where $S_k = \left\{\tilde{Q}(\tilde{Z},\alpha),\alpha\in\Lambda_k\right\}$, and $S^* = \bigcup_k S_k$.

Consider admissible structures-the structures that satisfy the following properties:

Any element S_k of structure $\mathbb{S}$ has a finite VC dimension $h_{\tilde{Z}}^k$.

Any element S_k of the structure (5.3) contains either

(i) a set of totally bounded measurable complex fuzzy set valued functions $\tilde{Q}(\tilde{Z},\alpha),\alpha\in\Lambda_k$, satisfying the following:

$A_{k1} \le \left\{\mathrm{Re}\left[\tilde{Q}(\tilde{Z},\alpha)\right]\right\}_r \le B_{k1}$ and $A_{k2} \le \left\{\mathrm{Im}\left[\tilde{Q}(\tilde{Z},\alpha)\right]\right\}_r \le B_{k2}, \forall r\in(0,1], \forall\alpha\in\Lambda_k$, where $A_{k1}, A_{k2}, B_{k1}, B_{k2} \in \mathbb{R}$. We also use the notation $A_k = \min\{A_{k1}, A_{k2}\}$ and $B_k = \max\{B_{k1}, B_{k2}\}$.

(ii) or a set of totally bounded nonnegative measurable complex fuzzy set valued functions $\tilde{Q}(\tilde{Z},\alpha),\alpha\in\Lambda_k$, satisfying the following:

$0 \le \left\{\mathrm{Re}\left[\tilde{Q}(\tilde{Z},\alpha)\right]\right\}_r \le B_{k1}$ and $0 \le \left\{\mathrm{Im}\left[\tilde{Q}(\tilde{Z},\alpha)\right]\right\}_r \le B_{k2}, \forall r\in(0,1], \forall\alpha\in\Lambda_k$, where $B_{k1}, B_{k2}\in\mathbb{R}$. We also use the notation $B_k = \max\{B_{k1}, B_{k2}\}$.

3. The set S^* is everywhere dense in the set S $\left(S = \left\{\tilde{Q}(\tilde{Z},\alpha),\alpha\in\Lambda\right\}\right)$ in the Hausdorff metric d, that is, $\forall\tilde{Q}(\tilde{Z},\alpha)\in S, \forall\varepsilon>0$, there exists a function $\tilde{Q}(\tilde{Z},\alpha^*)\in S^*$ such that $\int d\left(\tilde{Q}(\tilde{Z},\alpha),\tilde{Q}(\tilde{Z},\alpha^*)\right)dP < \varepsilon$.

Note that in view of the structure (5.3) the following assertions are true:

l. The sequence of values of VC dimension $h_{\tilde{Z}}^k$ for the elements S_k of the structure $\mathbb{S}$ is nondecreasing with increasing k:

$$h_{\tilde{Z}}^1 \le h_{\tilde{Z}}^2 \le \cdots \le h_{\tilde{Z}}^n \le \cdots.$$

2. The sequence of values of the bounds B_k for the elements S_k of the structure $\mathbb{S}$ is nondecreasing with

increasing k :

$$B_1 \le B_2 \le \cdots \le B_n \le \cdots .$$

Denote by $\tilde{Q}\left(\tilde{Z}, \alpha_l^k\right)$ the function that minimizes the complex fuzzy empirical risk in the set of functions S_k . For a given set of observations $\tilde{Z}_1, \cdots, \tilde{Z}_l$ the SRM method chooses the element S_k of the structure for which the smallest bound on the risk (the smallest guaranteed risk) is achieved.

Therefore the idea of the structural risk minimization principle of complex fuzzy random samples is the following:

To provide the given set of functions with an admissible structure and then to find the function that minimizes guaranteed risk over given elements of the structure.

To stress the importance of choosing the element of the structure that possesses an appropriate capacity, we call this principle the complex fuzzy structural risk minimization principle of statistical learning theory, the CFSRM principle, to be brief.

The CFSRM principle makes a compromise between the accuracy of approximation of the training data and the complexity on the set of approximated functions. The complex fuzzy empirical risk is decreased with the increased of the index of element of the structure, while the confidence interval is increased. The smallest bound of the risk is achieved on some appropriate element of the structure.

6. Consistency of the complex fuzzy structural risk minimization principle and asymptotic bounds on the rate of convergence

In this section we analyze asymptotic properties of the CFSRM principle. Here we answer two questions:

Is the CFSRM principle consistent? Do the risks for the functions chosen according to this principle converge to the smallest possible risk for the set S with increasing amount of observations?what is the bound on the (asymptotic) rate of convergence?

Let S be a set of measurable complex fuzzy set valued functions and let $\mathbb{S}$ be an admissible structure. Consider now the case where the structure contains an infinite number of elements. We denote by $\tilde{Q}\left(\tilde{Z}, \alpha_l^k\right), k = 1, \cdots$ the measurable complex fuzzy set valued function which minimizes the complex fuzzy empirical risk over the functions in the set S_k and denote by $\tilde{Q}\left(\tilde{Z}, \alpha_0^k\right)$ the measurable complex fuzzy set valued function which minimizes the complex fuzzy expected risk over the functions in the set $S_k s$; we denote also by $\tilde{Q}\left(\tilde{Z}, \alpha_0\right)$ the measurable complex fuzzy set valued function which minimizes the complex fuzzy expected risk over the set of function S . In the following text, we prove the consistency of the CFSRM principle. Consider the a priori rule $n = n(l)$ for choosing the number of element of the structure depending on the number of given samples.

Theorem 6.1. *The rule* $n = n(l)$ *provides approximations* $\tilde{Q}\left(\tilde{Z}, \alpha_l^{n(l)}\right)$ *for which the sequence of risks* $\tilde{R}_{cf}\left(\alpha_l^{n(l)}\right)$ *converges, as* l *tends to infinity, to the smallest risk:*

$$\tilde{R}_{cf}\left(\alpha_0\right) = \inf_{\alpha \in \Lambda} \tilde{R}_{cf}\left(\alpha\right)$$

with asymptotic rate of convergence

$$V(l) = r_{n(l)} + \sqrt{\frac{B_{n(l)}^2 h_{\tilde{Z}}^{n(l)} \ln l}{l}}, \tag{6.1}$$

where

$$r_{n(l)} = \left\|\tilde{R}_{cf}\left(\alpha_0^{n(l)}\right)\right\| - \left\|\tilde{R}_{cf}\left(\alpha_0\right)\right\|, \tag{6.2}$$

(that is, the equality $P\left\{\lim_{l\to\infty}\sup V^{-1}(l)\left\|\left\|\tilde{R}_{cf}\left(\alpha_l^{n(l)}\right)\right\| - \left\|\tilde{R}_{cf}\left(\alpha_0\right)\right\|\right\| < \infty\right\} = 1$ holds true), if

$$\frac{B_{n(l)}^2 h_{\tilde{Z}}^{n(l)} \ln l}{l} \xrightarrow[l\to\infty]{} 0, n(l) \xrightarrow[l\to\infty]{} \infty . \tag{6.3}$$

Proof: Consider a structure with elements S_k containing totally bounded functions with the finite VC dimension. We have that the following inequality

$$\Delta(\alpha_l) = \left\|\tilde{R}_{cf}(\alpha_l)\right\| - \left\|\tilde{R}_{cf}(\alpha_0)\right\| \le (B-A)\sqrt{4\frac{h_{\tilde{Z}}\left(\ln\frac{2l}{h_{\tilde{Z}}}+1\right)-\ln\left(\frac{\eta}{8}\right)}{l}} + (B-A)\sqrt{-\frac{2\ln\left(\frac{\eta}{4}\right)}{l}}$$

holds true with probability at least $1-2\eta$. Let $1-2\eta = 1-\frac{2}{l^2}$, that is, $\eta = \frac{1}{l^2}$. Then with probability $1-\frac{2}{l^2}$ the inequality

$$\Delta(\alpha_l) = \left\|\tilde{R}_{cf}(\alpha_l)\right\| - \left\|\tilde{R}_{cf}(\alpha_0)\right\| \le (B-A)\sqrt{4\frac{h_{\tilde{Z}}\left(\ln\frac{2l}{h_{\tilde{Z}}}+1\right)+\ln\left(8l^2\right)}{l}} + (B-A)\sqrt{\frac{4\ln(2l)}{l}}$$

is valid.

For any elements S_k with probability at least $1-\frac{2}{l^2}$ the additive bound

$$\Delta\left(\alpha_l^k\right) = \left\|\tilde{R}_{cf}\left(\alpha_l^k\right)\right\| - \left\|\tilde{R}_{cf}\left(\alpha_0^k\right)\right\| \le (B_k - A_k)\left(\sqrt{4\frac{h_{\tilde{Z}}^k\left(\ln\frac{2l}{h_{\tilde{Z}}^k}+1\right)+\ln\left(8l^2\right)}{l}} + \sqrt{\frac{4\ln(2l)}{l}}\right)$$

is valid. Then with probability $1-\frac{2}{l^2}$ the inequality

$$\left\|\tilde{R}_{cf}\left(\alpha_l^{n(l)}\right)\right\| - \left\|\tilde{R}_{cf}(\alpha_0)\right\| \le r_{n(l)} + \left(B_{n(l)} - A_{n(l)}\right)\left(\sqrt{4\frac{h_{\tilde{Z}}^{n(l)}\left(\ln\frac{2l}{h_{\tilde{Z}}^{n(l)}}+1\right)+\ln\left(8l^2\right)}{l}} + \sqrt{\frac{4\ln(2l)}{l}}\right) \tag{6.4}$$

holds, where

$$r_{n(l)} = \left\|\tilde{R}_{cf}\left(\alpha_0^{n(l)}\right)\right\| - \left\|\tilde{R}_{cf}(\alpha_0)\right\|.$$

Since $S^* = \bigcup_k S_k$ everywhere dense in $S = \left\{\tilde{Q}\left(\tilde{Z},\alpha\right), \alpha\in\Lambda\right\}$, for $\tilde{Q}\left(\tilde{Z},\alpha_0\right)\in S$, for any $\varepsilon > 0$, there exists a positive integer K such that $\tilde{Q}\left(\tilde{Z},\alpha^*\right)\in S_K$ and $\int d\left(\tilde{Q}\left(\tilde{Z},\alpha_0\right),\tilde{Q}\left(\tilde{Z},\alpha^*\right)\right)dP < \varepsilon$. We have

$$\begin{aligned}
&d\left(\tilde{R}_{cf}(\alpha^*), \tilde{R}_{cf}(\alpha_0)\right)\\
&= d\left(\int \mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha^*\right)\right]dP + i\int \mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha^*\right)\right]dP, \int \mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha_0\right)\right]dP + i\int \mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha_0\right)\right]dP\right)\\
&\le d\left(\int \mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha^*\right)\right]dP, \int \mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha_0\right)\right]dP\right) + d\left(\int \mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha^*\right)\right]dP, \int \mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha_0\right)\right]dP\right).
\end{aligned}$$

According to Theorem 3.3 of Ref. [17], we obtain

$$\begin{aligned}
&d\left(\int \mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha^*\right)\right]dP, \int \mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha_0\right)\right]dP\right) + d\left(\int \mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha^*\right)\right]dP, \int \mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha_0\right)\right]dP\right)\\
&\le \int d\left(\mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha^*\right)\right], \mathrm{Re}\left[\tilde{Q}\left(\tilde{Z},\alpha_0\right)\right]\right)dP + \int d\left(\mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha^*\right)\right], \mathrm{Im}\left[\tilde{Q}\left(\tilde{Z},\alpha_0\right)\right]\right)dP\\
&\le \int d\left(\tilde{Q}\left(\tilde{Z},\alpha_0\right), \tilde{Q}\left(\tilde{Z},\alpha^*\right)\right)dP + \int d\left(\tilde{Q}\left(\tilde{Z},\alpha_0\right), \tilde{Q}\left(\tilde{Z},\alpha^*\right)\right)dP \le 2\varepsilon.
\end{aligned}$$

When $n(l) \ge K$, the following relationships

$$\lim_{l\to\infty} r_{n(l)} = \lim_{l\to\infty}\left(\left\|\tilde{R}_{cf}\left(\alpha_0^{n(l)}\right)\right\| - \left\|\tilde{R}_{cf}\left(\alpha_0\right)\right\|\right) \le \lim_{l\to\infty}\left(\left\|\tilde{R}_{cf}\left(\alpha_0^{K}\right)\right\| - \left\|\tilde{R}_{cf}\left(\alpha_0\right)\right\|\right).$$

$$\le \lim_{l\to\infty}\left(\left\|\tilde{R}_{cf}\left(\alpha^{*}\right)\right\| - \left\|\tilde{R}_{cf}\left(\alpha_0\right)\right\|\right) \le \lim_{l\to\infty} d\left(\tilde{R}_{cf}\left(\alpha^{*}\right), \tilde{R}_{cf}\left(\alpha_0\right)\right) = 0$$

are satisfied.

Therefore the condition

$$\frac{B_{n(l)}^2 h_{\tilde{Z}}^{n(l)} \ln l}{l} \xrightarrow[l\to\infty]{} 0$$

determines convergence to zero. Denote

$$V(l) = r_{n(l)} + \left(B_{n(l)} - A_{n(l)}\right)\left(\sqrt{4\frac{h_{\tilde{Z}}^{n(l)}\left(\ln\frac{2l}{h_{\tilde{Z}}^{n(l)}} + 1\right) + \ln\left(8l^2\right)}{l}} + \sqrt{\frac{4\ln(2l)}{l}}\right).$$

Let we rewrite the assertion (6.4) in the form

$$P\left\{V^{-1}(l)\left(\left\|\tilde{R}_{cf}\left(\alpha_l^{n(l)}\right)\right\| - \left\|\tilde{R}_{cf}\left(\alpha_0\right)\right\|\right) > 1\right\} < \frac{2}{l^2}, l > l_0.$$

Since

$$\sum_{l=1}^{\infty} P\left\{V^{-1}(l)\left(\left\|\tilde{R}_{cf}\left(\alpha_l^{n(l)}\right)\right\| - \left\|\tilde{R}_{cf}\left(\alpha_0\right)\right\|\right) > 1\right\} < l_0 + \sum_{l=l_0+1}^{\infty} \frac{2}{l^2} < \infty$$

according to the corollary from the Borel-Cantelli lemma (see Ref. [28]), one can assert that the inequality

$$\overline{\lim_{l\to\infty}} V^{-1}(l)\left(\left\|\tilde{R}_{cf}\left(\alpha_l^{n(l)}\right)\right\| - \left\|\tilde{R}_{cf}\left(\alpha_0\right)\right\|\right) \le 1$$

is valid with probability one.

The next theorem is devoted to asymptotic properties of the complex fuzzy structural risk minimization principle.

Theorem 6.2. If the structure is such that $B_n^2 \le n^{1-\delta}$, then for any distribution function the CFSRM method provides convergence to the best possible solution with probability one (i.e., the CFSRM method is universally strongly consistent) . Moreover, if the optional solution $\tilde{Q}\left(\tilde{Z}, \alpha_0\right)$ belongs to some element S_*, of the structure $\left(\tilde{Q}\left(\tilde{Z}, \alpha_0\right) = \tilde{Q}\left(\tilde{Z}, \alpha^*\right)\right)$ and $B_{n(l)}^2 \le \mu(l) \le l^{1-\delta}$, then using the CFSRM method one achieves the following asymptotic rate of convergence: $V(l) = O\left(\sqrt{\frac{\mu(l)\ln l}{l}}\right)$.

Proof. To avoid choosing the minimum of functional (5.1) over the infinite number of elements of the structure, we introduce one additional constraint on the CFSRM method: we will choose the minimum from the first l elements of the structure where l is equal to the number of observations. Therefore we approximate the solution by function $\tilde{Q}\left(\tilde{Z}, \alpha_l^{+}\right)$, which among l functions $\tilde{Q}\left(\tilde{Z}, \alpha_l^{k}\right), k = 1, 2, \cdots, l$, minimizing empirical risk on corresponding elements $S_k, k = 1, 2, \cdots, l$, of the structure provide the smallest guaranteed (with probability $1 - \frac{1}{l}$) risk:

$$\left\|\tilde{R}^{+}_{cfemp}\left(\alpha_l^{+}\right)\right\| = \min_{1\le k\le l}\left[\left\|\tilde{R}_{cfemp}\left(\alpha_l^{k}\right)\right\| + \left(B_k - A_k\right)\sqrt{4\frac{h^k_{\tilde{Z}}\left(\ln\frac{2l}{h^k_{\tilde{Z}}}+1\right)+\ln\left(8l\right)}{l}}\right].$$

Denote by α_l^{+} the parameter that minimizes guaranteed risk $\tilde{R}^{+}_{cfemp}\left(\alpha\right)$ using l observations. Consider the decomposition

$$\left\|\tilde{R}_{cf}\left(\alpha_l^{+}\right)\right\| - \left\|\tilde{R}_{cf}\left(\alpha_0\right)\right\| = \left(\left\|\tilde{R}_{cf}\left(\alpha_l^{+}\right)\right\| - \left\|R^{+}_{cfemp}\left(\alpha_l^{+}\right)\right\|\right) + \left(\left\|R^{+}_{cfemp}\left(\alpha_l^{+}\right)\right\| - \left\|\tilde{R}_{cf}\left(\alpha_0\right)\right\|\right).$$

For the first term of this decomposition we have

$$P\left\{\left\|\tilde{R}_{cf}\left(\alpha_l^{+}\right)\right\| - \left\|R^{+}_{cfemp}\left(\alpha_l^{+}\right)\right\| > \varepsilon\right\} < \sum_{k=1}^{l} P\left\{\left\|\tilde{R}_{cf}\left(\alpha_l^{k}\right)\right\| - \left\|R^{+}_{cfemp}\left(\alpha_l^{k}\right)\right\| > \varepsilon\right\}$$

$$= \sum_{k=1}^{l} P\left\{\left\|\tilde{R}_{cf}\left(\alpha_l^{k}\right)\right\| - \left\|\tilde{R}_{cfemp}\left(\alpha_l^{k}\right)\right\| > \varepsilon + B_k\sqrt{4\frac{h^k_{\tilde{Z}}\left(\ln\frac{2l}{h^k_{\tilde{Z}}}+1\right)+\ln\left(8l\right)}{l}}\right\}$$

$$\le \sum_{k=1}^{l} 4\exp\left\{-\frac{l}{2}\left(\frac{\varepsilon}{B_k} + \sqrt{4\frac{h^k_{\tilde{Z}}\left(\ln\frac{2l}{h^k_{\tilde{Z}}}+1\right)+\ln\left(8l\right)}{l}}\right)^2\right\}$$

$$\le \sum_{k=1}^{l} 32\left(\frac{2le}{h^k_{\tilde{Z}}}\right)^{h^k_{\tilde{Z}}} \exp\left\{-\frac{l}{4}\left(\frac{\varepsilon}{B_k} + \sqrt{4\frac{h^k_{\tilde{Z}}\left(\ln\frac{2l}{h^k_{\tilde{Z}}}+1\right)+\ln\left(8l\right)}{l}}\right)^2\right\}$$

$$\le \sum_{k=1}^{l}\frac{4}{l}\exp\left\{-\frac{l\varepsilon^2}{4B_k^2}\right\} \le 4\exp\left\{-\frac{l\varepsilon^2}{4B_l^2}\right\} < 4\exp\left\{-\frac{\varepsilon^2 l^{\delta}}{4}\right\}$$

where we take into account that $B_l^2 = l^{1-\delta}$. Using the Borel-Cantelli lemma we obtain that first summand of the decomposition converges almost surely to the non-positive value.

Now consider the second term of the decomposition. Since $S^{*} = \bigcup_k S_k$ everywhere dense in $S = \left\{\tilde{Q}\left(\tilde{Z},\alpha\right), \alpha\in\Lambda\right\}$, for every ε there exists an elements S_t of the structure such that

$$\left\|\tilde{R}_{cf}\left(\alpha_0^{t}\right)\right\| - \left\|\tilde{R}_{cf}\left(\alpha_0\right)\right\| \le \varepsilon$$

Therefore we will prove that the second term in the decomposition does not exceed zero if we show that with probability one

$$\lim_{l\to\infty}\min_{1\le k\le l}\left(\left\|\tilde{R}^{+}_{cfemp}\left(\alpha_l^{k}\right)\right\| - \left\|\tilde{R}_{cf}\left(\alpha_0^{t}\right)\right\|\right) \le 0$$

Note that for any ε there exists l_0 such that for any $l > l_0$

$$B_t\sqrt{4\frac{h^t_{\tilde{Z}}\left(\ln\frac{2l}{h^t_{\tilde{Z}}}+1\right)+\ln\left(8l\right)}{l}} \le \frac{\varepsilon}{2}. \tag{6.5}$$

For $l > l_0$ we have

$$P\left\{\min_{1\le k\le l}\left\|\tilde{R}^{+}_{cfemp}\left(\alpha_l^k\right)\right\|-\left\|\tilde{R}_{cf}\left(\alpha_0^t\right)\right\|>\varepsilon\right\}\le P\left\{\left\|\tilde{R}^{+}_{cfemp}\left(\alpha_l^t\right)\right\|-\left\|\tilde{R}_{cf}\left(\alpha_0^t\right)\right\|>\varepsilon\right\}$$

$$=P\left\{\left\|\tilde{R}_{cfemp}\left(\alpha_l^t\right)\right\|-\left\|\tilde{R}_{cf}\left(\alpha_0^t\right)\right\|>\varepsilon-B_t\sqrt{4\frac{h^t_{\tilde{Z}}\left(\ln\frac{2l}{h^t_{\tilde{Z}}}+1\right)+\ln(8l)}{l}}\right\}$$

$$\le P\left\{\left\|\tilde{R}_{cfemp}\left(\alpha_l^t\right)\right\|-\left\|\tilde{R}_{cf}\left(\alpha_0^t\right)\right\|>\frac{\varepsilon}{2}\right\}\le P\left\{\sup_{\alpha\in\Lambda_t}\left\|\tilde{R}_{cf}\left(\alpha\right)\right\|-\left\|\tilde{R}_{cfemp}\left(\alpha\right)\right\|>\frac{\varepsilon}{2}\right\}$$

$$\le 16\exp\left\{\left(\frac{h^t_{\tilde{Z}}\left(\ln\frac{2l}{h^t_{\tilde{Z}}}+1\right)}{l}-\frac{\varepsilon^2}{8B_t^2}\right)l\right\}\le 16\left(\frac{2le}{h^t_{\tilde{Z}}}\right)^{h^t_{\tilde{Z}}}\exp\left\{-\frac{\varepsilon^2 l}{8B_t^2}\right\}<16\left(\frac{2le}{h^t_{\tilde{Z}}}\right)^{h^t_{\tilde{Z}}}\exp\left\{-\frac{\varepsilon^2 l^{\delta}}{8}\right\}.$$

Again applying the Borel-Cantelli lemma one concludes that second term of the decomposition converges almost surely to the nonpositive value. Since the sum of two terms is nonnegative, we obtain almost sure convergence $\left\|\tilde{R}_{cf}\left(\alpha_l^+\right)\right\|$ to $\left\|\tilde{R}_{cf}\left(\alpha_0\right)\right\|$. This proves the first part of the theorem.

To prove the second part, note that when the optimal solution belongs to one of the elements of the structure S_t the equality $\left\|\tilde{R}_{cf}\left(\alpha_0^t\right)\right\|=\left\|\tilde{R}_{cf}\left(\alpha_0\right)\right\|$ holds true. Combining bounds for both terms, one obtains that for l satisfying (6.5) the following inequalities are valid:

$$P\left\{\left\|\tilde{R}_{cf}\left(\alpha_l^+\right)\right\|-\left\|\tilde{R}_{cf}\left(\alpha_0\right)\right\|>\varepsilon\right\}$$

$$\le P\left\{\left\|\tilde{R}_{cf}\left(\alpha_l^+\right)\right\|-\left\|R^{+}_{cfemp}\left(\alpha_l^+\right)\right\|>\frac{\varepsilon}{2}\right\}+P\left\{\left\|R^{+}_{cfemp}\left(\alpha_l^+\right)\right\|-\left\|\tilde{R}_{cf}\left(\alpha_0\right)\right\|>\frac{\varepsilon}{2}\right\}$$

$$\le 4\exp\left\{-\frac{\varepsilon^2 l}{16\mu(l)}\right\}+16\left(\frac{2le}{h^t_{\tilde{Z}}}\right)^{h^t_{\tilde{Z}}}\exp\left\{-\frac{\varepsilon^2 l}{32\mu(l)}\right\}.$$

Form this inequality we obtain the rate of convergence:

$$V(l)=O\left(\sqrt{\frac{\mu(l)\ln l}{l}}\right).$$

7. Conclusions

Considering the existence and the significance of complex fuzzy random variables in real world, this paper proposes the concepts of capacity of the set of measurable complex fuzzy set valued functions and the structural risk minimization principle based on complex fuzzy random samples. Furthermore, the consistency of the complex fuzzy structural risk minimization principle is proven, and bound on the asymptotic rate of convergence is presented. Altogether these findings have laid the foundation for further research in statistical learning theory involving complex fuzzy random samples. Further investigations might focus on some applied aspects such as e.g., complex fuzzy support vector machines.

8. Acknowledgements

The author would like to thank the referees for their valuable comments and suggestions.

9. References

[1] M. Brunato, R. Battiti. Statistical learning theory for location fingerprinting in wireless LANs. *Computer Networks.* 2005, **47**: 825–845.

[2] J. J. Buckley. Fuzzy complex numbers. *Fuzzy Sets and Systems.* 1989, **33**: 333-345.

[3] J. J. Buckley. Fuzzy complex analysis II: integration. *Fuzzy Sets and Systems*. 1992, **49**: 171-179.

[4] J. J. Buckley, Y. X. Qu. Fuzzy complex analysis I: differentiation. *Fuzzy Sets and Systems.* 1991, **41**: 269-284.

[5] J. L. Castro, L. D. Flores-Hidalgo, C. J. Mantas, J. M. Puche. Extraction of fuzzy rules from support vector machines. *Fuzzy Sets and Systems*. 2007, **158**: 2057-2077.

[6] A. Celikyilmaz, I. B. Turksen. Fuzzy functions with support vector machines. *Information Sciences*. 2007, **177**: 5163-5177.

[7] T. Evgeniou, T. Poggio, M. Pontil, et al. Regularization and statistical learning theory for data analysis. *Computational Statistics & Data Analysis*. 2002, **38**: 421-432.

[8] M. H. Ha, Y. C. Bai, P. Wang, et al. The key theorem and the bounds on the rate of uniform convergence of statistical learning theory on a credibility space. *Advances in Fuzzy Sets and Systems*. 2006, **1**: 143-172.

[9] M. H. Ha, Z. F. Feng, S. J. Song, L. Q. Gao. The key theorem and the bounds on the rate of uniform convergence of statistical learning theory on quasi-probability spaces. *Chinese Journal of Computers*. 2008, **31**: 476-485.

[10] M. H. Ha, Y. Li, J. Li, D. Z. Tian. The key theorem and the bounds on the rate of uniform convergence of learning theory on Sugeno measure space. *Science in China*: *Series F* . 2006, **49**: 1-14.

[11] M. H. Ha, J. Tian. The theoretical foundations of statistical learning theory based on fuzzy number samples. *Information Sciences*. 2008, 178: 3240-3246.

[12] C. L. Huang, M.C. Chen, C.J. Wang. Credit scoring with a data mining approach based on support vector machines. *Expert System with Applications*. 2007, **33**: 847-856.

[13] C. Hwang, D. H. Hong, K. H. Seok. Support vector interval regression machine for crisp input and output data. *Fuzzy Sets and Systems*. 2006, **157**: 1114-1125.

[14] B. Jin, Y. C. Tang, Y. Q. Zhang. Support vector machines with genetic fuzzy feature transformation for biomedical data classification. *Information Sciences*. 2007, **177**: 476-489.

[15] T. Kikuchi, S. Abe. Comparison between error correcting output codes and fuzzy support vector machines. *Pattern Recognition Letters*. 2005, **26**: 1937-1945.

[16] Y. K. Kim. A strong law of large numbers for fuzzy random variables. *Fuzzy Sets and Systems*. 2000, **111**: 319-323.

[17] Y. K. Kim, B. M. Ghil. Integrals of fuzzy-number-valued functions. *Fuzzy Sets and Systems*. 1997, **86**: 213-222.

[18] C. F. Lin, S. D. Wang. Fuzzy support vector machines. *IEEE Transactions on Neural Networks*. 2002, **13**: 464-471.

[19] B. D. Liu, Y. K. Liu. Expected value of fuzzy variable and fuzzy expected value models. *IEEE Transactions on Fuzzy Systems.* 2002, **10**: 445-450.

[20] Y. H. Liu, Y. T. Chen. Face recognition using total margin-based adaptive fuzzy support vector machines. *IEEE Transactions on Neural Networks.* 2007, **18**: 178-192.

[21] D. Qiu, L. Shu. Notes on "On the restudy of fuzzy complex analysis: Part I and Part II". *Fuzzy Sets and Systems*. 2008, **159**: 2185-2189.

[22] J. Q. Qiu, C. X. Wu, F. C. Li. On the restudy of fuzzy complex analysis: Part I. The sequence and series of fuzzy complex numbers and their convergences. *Fuzzy Sets and Systems*. 2000, **115**: 445-450.

[23] J. Q. Qiu, C. X. Wu, F. C. Li. On the restudy of fuzzy complex analysis: Part II. The continuity and differentiation of fuzzy complex functions. *Fuzzy Sets and Systems*. 2001, **120**: 517-521.

[24] C. Su, C. H. Yang. Feature selection for the SVM: an application to hypertension diagnosis. *Expert Systems with Applications*. 2008, **34**: 754-763.

[25] D. Z. Tian, Z. M. Zhang, M. H. Ha. The key theorem of complex statistical learning theory. *Dynamics of Continuous, Discrete and Impulsive Systems, Series A: Mathematical Analysis.* 2007, **14**(S1): 46-50.

[26] J. J. Tsong. Hybrid approach of selecting hyper-parameters of support vector machine for regression. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*. 2006, **36**: 699-709.

[27] V. N. Vapnik. *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[28] V. N. Vapnik. *Statistical Learning Theory*. New York: A Wiley-Interscience Publication, 1998.

[29] V. N. Vapnik. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*. 1999, **10**: 988-999.

[30] M. Vidyasagar. Statistical learning theory and randomized algorithms for control. *IEEE Control Systems Magazine.* 1998, **12**: 69-85.

[31] G. J. Wang, X. P. Li. Generalized Lebesgue integrals of fuzzy complex valued functions. *Fuzzy Sets and Systems.* 2002, **127**: 363-370.

[32] Y. Q. Wang, S. Y. Wang, K. K. Lai. A new fuzzy support vector machine to evaluate credit risk. *IEEE Transactions on Fuzzy Systems*. 2005, **13**: 820-931.

[33] Z. Y. Wang, J. K. George. *Fuzzy Measure Theory*. New York: Plenum Press, 1992.

[34] H. Wechsler, Z. Duric, F. Y. Li, V.. Cherkassky, Motion estimation using statistical learning theory. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2004, **26**: 466-478.

[35] C. X. Wu, J. Q. Qiu. Some remarks for fuzzy complex analysis. *Fuzzy Sets and Systems*. 1999, **106**: 231-238.

[36] S. J. Yan, J. X. Wang, Q. X. Liu. *Foundations of Probability Theory*. Beijing: Science Press, 1999.

[37] G. Q. Zhang. Fuzzy limit theory of fuzzy complex numbers. *Fuzzy Sets and Systems*. 1998, **46**: 227-235.

[38] Y. Zhang, G. Y. Wang. *Fuzzy Random Power System Theory*. Beijing: Science Press, 1993.