# Discrimination for Fuzzy Sets related to NTV Metric

Guoxiang Lu, Rui Liu

***Abstract*—In this paper, we study a discrimination for the fuzzy sets based on the averaging operators. The discrimination generalizes the known NTV metric by a free parameter $\gamma \in [0,1]$. We give the metric properties of the discrimination with respect to the parameter $\gamma$, and obtain the sufficient and necessary condition when the discrimination is a metric. Furthermore, we show that the discrimination is always a metric, called RNTV metric, on the positive parts of spheres with their centers under $l_1$-norm. Finally, by computing basic examples of codons, we show some numerical comparison for the new RNTV metric to original NTV metric.**

***Index Terms*—Discrimination, Fuzzy set, NTV metric, Triangle inequality, RNTV metric.**

## I. Introduction

Sequence analysis and sequence comparison have become two fundamental methods in the modern molecular biology. In the past time, many pieces of research were made to obtain more information about the sequences. The structure comparison algorithms for molecular sequences has been discussed in [1], [2]. To study the similarities and dissimilarities of genetic sequences, a new metric was introduced by Nieto, Torres and Vázquez-Trasande ([3]). This metric is based on the idea of fuzzy Hamming distance and fuzzy entropy theorem ([4], [5], [6]). It plays a very important role in mathematical biology, especially in sequence analysis and sequence comparison ([7], [8]). In [9], Dress and Lokot called it NTV metric, and presented a simple proof of the triangle inequality for the new metric.

In section 2, we introduce the relationship between NTV metric and fuzzy operators. Then by using the concept in [10], [11], a discrimination related to the NTV metric is presented. There exists a free parameter $\gamma \in [0,1]$ in the new discrimination. In section 3, we discuss the new discrimination with its parameter $\gamma$ and give its metric properties. Moreover, we obtain the sufficient and necessary condition when the discrimination is a metric. Finally, we show that the discrimination is always a metric, called RNTV metric, on the positive parts of spheres with their centers under $l_1$-norm. In section 4, we numerically compare the new RNTV metric to the NTV metric by computing some basic examples of codons.

Manuscript received September 3, 2013; revised November 20, 2013.

This work was supported by the National Natural Science Foundation of China (the NSF of China) Grant No. 11001134 and 11201336, and the Fundamental Research Funds for the Central Universities (FRF-CU) Grant No. 2722013JC082.

Guoxiang Lu is with School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan 430073, China. E-mail: lgxmath@znufe.edu.cn

Rui Liu is with Department of Mathematics and LPMC, Nankai University, Tianjin 300071, China. Email: ruiliu@nankai.edu.cn

Corresponding author: E-mail: ruiliu@nankai.edu.cn

## II. Preliminaries and Discrimination related to NTV Metric

Let $X = \{x_1, x_2, \cdots, x_n\}$ be a fixed set, a fuzzy set in $X$ is defined by

$$A = \{(x, \mu_A(x)) | \, x \in X\}$$

where

$$\mu_A : X \to I = [0,1], \, x \mapsto \mu_A(x).$$

The number $\mu_A(x)$ denotes the membership degree of the element $x$ in the fuzzy set $A$. We can also use the unit hypercube $I^n = [0,1]^n$ to describe all the fuzzy sets in $X$, because a fuzzy set $A$ determines a point $P = (\mu_A(x_1), \mu_A(x_2), \cdots, \mu_A(x_n))$. Reciprocally, any point $P = (a_1, a_2, \cdots, a_n) \in I^n$ generates a fuzzy set $A$ defined by $\mu_A(x_i) = a_i, i = 1, 2, \cdots, n$.

Given two fuzzy sets $P = (p_1, p_2, \cdots, p_n)$, $Q = (q_1, q_2, \cdots, q_n) \in I^n$, it is defined Zadeh operators $\wedge$ and $\vee$ as

$$p_i \wedge q_i = \min\{p_i, q_i\}, \; p_i \vee q_i = \max\{p_i, q_i\}. \tag{1}$$

Consider the Zadeh operators $\wedge$ and $\vee$ for intersection and union of fuzzy sets $P$ and $Q$. If $P, Q$ are not both equal to zero point $\mathbf{0} := (0, 0, \cdots, 0)$, we can defined the similarity of $P$ and $Q$ as [6], [12]

$$\mathrm{Similar}(P,Q) = \frac{P \bigcap Q}{P \bigcup Q} = \frac{\sum_{i=1}^{n} (p_i \wedge q_i)}{\sum_{i=1}^{n} (p_i \vee q_i)}. \tag{2}$$

Of course, if $P = Q = \mathbf{0}$, then we define $\mathrm{Similar}(P,Q) = 1$.

It is easy to see $0 \le \mathrm{Similar}(P,Q) \le 1$. So we defined the difference of $P$ and $Q$ as [6]

$$\mathrm{Differ}(P,Q) = 1 - \mathrm{Similar}(P,Q). \tag{3}$$

Take (1) and (2) into (3),

$$\mathrm{Differ}(P,Q) = \frac{\sum_{i=1}^{n} \max\{p_i, q_i\} - \sum_{i=1}^{n} \min\{p_i, q_i\}}{\sum_{i=1}^{n} \max\{p_i, q_i\}}. \tag{4}$$

Using the relation

$$\sum_{i=1}^{n} |p_i - q_i| = \sum_{i=1}^{n} (\max\{p_i, q_i\} - \min\{p_i, q_i\}) \tag{5}$$

we have

$$\text{Differ}(P,Q) = \frac{\sum_{i=1}^{n} |p_i - q_i|}{\sum_{i=1}^{n} \max\{p_i, q_i\}}. \tag{6}$$

In [3], the NTV metric is defined by

$$d_{\text{NTV}}(P,Q) = \frac{\sum_{i=1}^{n} |p_i - q_i|}{\sum_{i=1}^{n} \max\{p_i, q_i\}}. \tag{7}$$

This shows that the formula (6) is just the definition of the NTV metric. So the NTV metric is the difference of fuzzy sets by intersection and union.

As the authors consider the degree of similarity related to the canonical midpoint between $P$ and $Q$ ([7]), we use the averaging operators $\triangle$ and $\triangledown$ extended from the concepts of $\wedge$ and $\vee$ ([10], [11]). A parameter $\gamma$ is added in their definitions as

$$p_i \triangle q_i = \gamma \min\{p_i, q_i\} + (1-\gamma)\frac{p_i + q_i}{2}, \tag{8}$$

$$p_i \triangledown q_i = \gamma \max\{p_i, q_i\} + (1-\gamma)\frac{p_i + q_i}{2}, \tag{9}$$

where $\gamma \in [0,1]$. These averaging operators actually combine the $\wedge$ and $\vee$ operators, respectively, with the arithmetic mean. Take (5), (8) and (9) into (2) and (3),

$$\begin{aligned}\text{sim}(P,Q) :=& \text{Similar}(P,Q)\\ =& \frac{2\gamma \sum_{i=1}^{n} \min\{p_i, q_i\} + (1-\gamma)\sum_{i=1}^{n}(p_i + q_i)}{2\gamma \sum_{i=1}^{n} \max\{p_i, q_i\} + (1-\gamma)\sum_{i=1}^{n}(p_i + q_i)},\end{aligned} \tag{10}$$

$$\begin{aligned}\text{dis}(P,Q) :=& \text{Differ}(P,Q)\\ =& \frac{2\gamma \left(\sum_{i=1}^{n} |p_i - q_i|\right)}{2\gamma \sum_{i=1}^{n} \max\{p_i, q_i\} + (1-\gamma)\sum_{i=1}^{n}(p_i + q_i)}.\end{aligned} \tag{11}$$

The parameter $\gamma$ essentially determines the uncertainty degree of $\triangle$ and $\triangledown$ operators [11], [13]. If $\gamma = 1$, then the operators are completely crisp. If $\gamma = 0$, then the operators are completely uncertain because we cannot differentiate the binary relation Min from Max. Therefore, the similarity value of any two fuzzy sets is 1 (100% similarity). When $\gamma = 1$, the discrimination of fuzzy sets is the NTV metric. Naturally, there exist two problems to explain: Is the discrimination (11) with all values of parameter $\gamma$ a metric? What is the geometric meaning of parameter $\gamma$? In section 3, the answers are presented.

## III. Main Results

**THEOREM 1.** If $n = 1$, then $\text{dis}(P,Q)$ is a metric for arbitrary $\gamma \in [0,1]$.

**PROOF.** Let $P = p, Q = q, R = r \in I$. It is obvious that the properties of nonnegativity and symmetry holds. So we only consider the triangle inequality.

In order to omit the absolute value in (11), we consider the order of $p, q, r$. Without loss of generality, we assume $p \le q \le r$. Then

$$\begin{aligned}\text{dis}(P,Q) &= \frac{2\gamma(q-p)}{2\gamma q + (1-\gamma)(p+q)},\\ \text{dis}(Q,R) &= \frac{2\gamma(r-q)}{2\gamma r + (1-\gamma)(q+r)},\\ \text{dis}(P,R) &= \frac{2\gamma(r-p)}{2\gamma r + (1-\gamma)(p+r)}.\end{aligned}$$

As $r - q \le r - p, p + r \le q + r$, we have

$$\text{dis}(Q,R) \le \text{dis}(P,R).$$

So

$$\text{dis}(P,R) + \text{dis}(P,Q) \ge \text{dis}(Q,R).$$

Let

$$f(x) = \frac{2\gamma(x-p)}{2\gamma x + (1-\gamma)(p+x)}.$$

By differentiation,

$$f'(x) = \frac{4\gamma p}{[2\gamma x + (1-\gamma)(p+x)]^2} \ge 0,$$

hence $f$ is nondecreasing. Thus, $\text{dis}(P,R) = f(r) \ge f(q) = \text{dis}(P,Q)$. So

$$\text{dis}(P,R) + \text{dis}(Q,R) \ge \text{dis}(P,Q).$$

For $\gamma \in [0,1], 0 \le p \le q \le r \le 1$,

$$\begin{aligned}&\frac{1+\gamma}{2\gamma q + (1-\gamma)(p+q)} \ge \frac{1-\gamma}{2\gamma r + (1-\gamma)(q+r)}\\ \Rightarrow&\frac{(q-p)(r-q)(1+\gamma)}{[2\gamma q + (1-\gamma)(p+q)][2\gamma r + (1-\gamma)(p+r)]} \ge\\ &\frac{(q-p)(r-q)(1-\gamma)}{[2\gamma r + (1-\gamma)(q+r)][2\gamma r + (1-\gamma)(p+r)]}\\ \Rightarrow&(q-p)\left[\frac{1}{2\gamma q + (1-\gamma)(p+q)} - \frac{1}{2\gamma r + (1-\gamma)(p+r)}\right] \ge\\ &(r-q)\left[\frac{1}{2\gamma r + (1-\gamma)(p+r)} - \frac{1}{2\gamma r + (1-\gamma)(q+r)}\right]\\ \Rightarrow&\frac{2\gamma(q-p)}{2\gamma q + (1-\gamma)(p+q)} + \frac{2\gamma(r-q)}{2\gamma r + (1-\gamma)(q+r)} \ge\\ &\frac{2\gamma(q-p)}{2\gamma r + (1-\gamma)(p+r)} + \frac{2\gamma(r-q)}{2\gamma r + (1-\gamma)(p+r)}\\ \Rightarrow&\frac{2\gamma(q-p)}{2\gamma q + (1-\gamma)(p+q)} + \frac{2\gamma(r-p)}{2\gamma r + (1-\gamma)(q+r)} \ge\\ &\frac{2\gamma(r-p)}{2\gamma r + (1-\gamma)(p+r)}.\end{aligned}$$

That is $\text{dis}(P,Q) + \text{dis}(Q,R) \ge \text{dis}(P,R)$. Thereby, the triangle inequality holds. □

**THEOREM 2.** If $n \ge 2$, then $\text{dis}(P,Q)$ is a metric if and only if $\gamma \in \{0,1\}$.

**PROOF.** If $\gamma = 0$, then $\text{dis}(P,Q) = 0$; If $\gamma = 1$, then $\text{dis}(P,Q) = d_{\text{NTV}}(P,Q)$. So they are both metrics.

If $\gamma \in (0,1)$, we consider the three points:

$$P = (0.1, 0.2, 0, \cdots, 0),$$

$$Q = (0.2, 0.2, 0, \cdots, 0),$$

$$R = (0.2, 0.1, 0, \cdots, 0).$$

Then

$$\mathrm{dis}(P,R)=\frac{2\gamma}{3+\gamma},\ \mathrm{dis}(P,Q)=\frac{2\gamma}{7+\gamma},\ \mathrm{dis}(Q,R)=\frac{2\gamma}{7+\gamma},$$

$$\mathrm{dis}(P,Q)+\mathrm{dis}(Q,R)=\frac{4\gamma}{7+\gamma}<\frac{4\gamma}{6+2\gamma}=\mathrm{dis}(P,R).$$

That shows the triangle inequality does not hold. □

**THEOREM 3.** Let $n\geq 2$, $C\in[0,n]$, and

$$S_C=\Big\{(x_1,x_2,\cdots,x_n)\in I^n:\sum_{i=1}^{n}x_i=C\Big\}.$$

Then $\mathrm{dis}(P,Q)$ is a metric on $S_C\cup\{\mathbf{0}\}$.

**PROOF.** Let $P,Q,R\in S_C\cup\{\mathbf{0}\}$,

$$P=(p_1,p_2,\cdots,p_n),$$

$$Q=(q_1,q_2,\cdots,q_n),$$

$$R=(r_1,r_2,\cdots,r_n).$$

It is obvious that $\mathrm{dis}(P,Q)=\mathrm{dis}(Q,P)\geq 0$ for arbitrary $P,Q\in I^n$. If $\mathrm{dis}(P,Q)=0$, then $\sum\limits_{i=1}^{n}|p_i-q_i|=0$, thus $p_i=q_i, i=1,2,\cdots,n$ and $P=Q$.

Next, to prove the triangle inequality

$$\mathrm{dis}(P,R)+\mathrm{dis}(R,Q)\geq\mathrm{dis}(P,Q),$$

we have to consider four cases.

1. The first case is when $P,Q,R\in S_C$. Formula (11) can be simplified as

$$\mathrm{dis}(P,Q)=\frac{2\gamma\sum\limits_{i=1}^{n}|p_i-q_i|}{2\gamma\sum\limits_{i=1}^{n}\max\{p_i,q_i\}+2(1-\gamma)C}=\frac{\sum\limits_{i=1}^{n}|p_i-q_i|}{\sum\limits_{i=1}^{n}\max\{p_i,q_i\}+\frac{1-\gamma}{\gamma}C}$$

Apply the method in [9], putting

$$A=\sum_{i=1}^{n}|p_i-q_i|,B=\sum_{i=1}^{n}\max\{p_i,q_i\},$$

$$A'=\sum_{i=1}^{n}(|p_i-r_i|+|r_i-q_i|),B'=\sum_{i=1}^{n}\max\{p_i,q_i,r_i\}.$$

As $A\leq A'$, $\frac{1-\gamma}{\gamma}C>0$, and the result $AB'\leq B'A$ holds in [9], we have

$$AB'+\frac{A(1-\gamma)C}{\gamma}\leq B'A+\frac{A'(1-\gamma)C}{\gamma}.$$

Both sides divided by $\left(B+\frac{1-\gamma}{\gamma}C\right)\left(B'+\frac{1-\gamma}{\gamma}C\right)$, we find

$$\frac{A}{B+\frac{1-\gamma}{\gamma}C}\leq\frac{A'}{B'+\frac{1-\gamma}{\gamma}C}.$$

The above shows

$$\mathrm{dis}(P,Q)\leq\frac{\sum\limits_{i=1}^{n}(|p_i-r_i|+|r_i-q_i|)}{\sum\limits_{i=1}^{n}\max\{p_i,q_i,r_i\}+\frac{1-\gamma}{\gamma}C}.$$

With

$$\frac{\sum\limits_{i=1}^{n}(|p_i-r_i|+|r_i-q_i|)}{\sum\limits_{i=1}^{n}\max\{p_i,q_i,r_i\}+\frac{1-\gamma}{\gamma}C}$$

$$=\frac{\sum\limits_{i=1}^{n}|p_i-r_i|}{\sum\limits_{i=1}^{n}\max\{p_i,q_i,r_i\}+\frac{1-\gamma}{\gamma}C}+\frac{\sum\limits_{i=1}^{n}|r_i-q_i|}{\sum\limits_{i=1}^{n}\max\{p_i,q_i,r_i\}+\frac{1-\gamma}{\gamma}C}$$

$$\leq\frac{\sum\limits_{i=1}^{n}|p_i-r_i|}{\sum\limits_{i=1}^{n}\max\{p_i,r_i\}+\frac{1-\gamma}{\gamma}C}+\frac{\sum\limits_{i=1}^{n}|r_i-q_i|}{\sum\limits_{i=1}^{n}\max\{q_i,r_i\}+\frac{1-\gamma}{\gamma}C},$$

the triangle inequality $\mathrm{dis}(P,R)+\mathrm{dis}(R,Q)\geq\mathrm{dis}(P,Q)$ holds.

2. The second case is when $P,R\in S_C, Q=\mathbf{0}$. Then

$$\mathrm{dis}(P,Q)=\frac{2\gamma C}{2\gamma C+(1-\gamma)C}=\frac{2\gamma}{1+\gamma}=\mathrm{dis}(R,Q),\quad(12)$$

Formula (12) and $\mathrm{dis}(P,R)\geq 0$ lead

$$\mathrm{dis}(P,R)+\mathrm{dis}(R,Q)\geq\mathrm{dis}(P,Q).$$

3. The third case is there exists only one zero point in $P,R$. In this case, we have

$$\mathrm{dis}(P,R)=\frac{2\gamma C}{2\gamma C+(1-\gamma)C}=\frac{2\gamma}{1+\gamma}.$$

If $P=\mathbf{0},R\in S_C$, $Q\in S_C$, then we have

$$\mathrm{dis}(P,Q)=\frac{2\gamma C}{2\gamma C+(1-\gamma)C}=\frac{2\gamma}{1+\gamma}=\mathrm{dis}(P,R).$$

If $P=\mathbf{0},R\in S_C$, $Q=\mathbf{0}$, then

$$\mathrm{dis}(P,Q)=0.$$

If $P\in S_C,R=\mathbf{0},Q\in S_C$,

$$\mathrm{dis}(P,Q)=\frac{\sum\limits_{p_i\geq q_i}(p_i-q_i)+\sum\limits_{p_i<q_i}(q_i-p_i)}{\sum\limits_{p_i\geq q_i}p_i+\sum\limits_{p_i<q_i}q_i+\frac{1-\gamma}{\gamma}C}\leq\frac{\sum\limits_{p_i\geq q_i}p_i+\sum\limits_{p_i<q_i}q_i}{\sum\limits_{p_i\geq q_i}p_i+\sum\limits_{p_i<q_i}q_i+\frac{1-\gamma}{\gamma}C}.$$

Noting

$$\sum_{p_i\geq q_i}p_i+\sum_{p_i<q_i}q_i\leq 2C,$$

we have

$$0\leq\mathrm{dis}(P,Q)\leq\frac{2C}{2C+\frac{1-\gamma}{\gamma}C}=\frac{2\gamma}{1+\gamma}.\qquad(13)$$

then from (13) we have

$$\mathrm{dis}(P,R)\geq\mathrm{dis}(P,Q).$$

If $P\in S_C,R=\mathbf{0}$, $Q=\mathbf{0}$, then we have

$$\mathrm{dis}(P,Q)=\frac{2\gamma C}{2\gamma C+(1-\gamma)C}=\frac{2\gamma}{1+\gamma}=\mathrm{dis}(P,R).$$

All above lead $\mathrm{dis}(P,R)+\mathrm{dis}(R,Q)\geq\mathrm{dis}(P,Q)$.

4. The fourth case is when $P = R = \mathbf{0}$. We have $\mathrm{dis}(P,R) = 0$, $\mathrm{dis}(P,Q) = \mathrm{dis}(R,Q)$. That leads $\mathrm{dis}(P,R) + \mathrm{dis}(R,Q) \geq \mathrm{dis}(P,Q)$.

To sum up the four cases, we have proved the validity of the triangle inequality on $S_C \cup \{\mathbf{0}\}$. □

We define the new metric in theorem 3 as $d_{\mathrm{RNTV}}$:

$$d_{\mathrm{RNTV}}(P,Q) = \frac{2\gamma\left(\sum\limits_{i=1}^{n}|p_i - q_i|\right)}{2\gamma\sum\limits_{i=1}^{n}\max\{p_i,q_i\} + (1-\gamma)\sum\limits_{i=1}^{n}(p_i+q_i)} \tag{14}$$

where $P,Q \in S_C \cup \{\mathbf{0}\}$.

**COROLLARY 1.** $d_{\mathrm{RNTV}}(P,Q)$ is a metric on the probability space.

**COROLLARY 2.** If $\gamma = 0$, then $d_{\mathrm{RNTV}}(P,Q) = 0$ .

**COROLLARY 3.** If $P \neq \mathbf{0}$, then

$$d_{\mathrm{RNTV}}(P,\mathbf{0}) = \frac{2\gamma}{1+\gamma}.$$

**COROLLARY 4.** If $P \neq \mathbf{0}$, then

$$\gamma = \frac{d_{\mathrm{RNTV}}(P,\mathbf{0})}{2 - d_{\mathrm{RNTV}}(P,\mathbf{0})}.$$

**COROLLARY 5.** If $\gamma \geq 0$ and $P,Q \in S_C$, then $d_{\mathrm{RNTV}}(P,Q) \leq \gamma$. The equality holds if and only if $p_i q_i = 0$ for all $i = 1,2,\cdots,n$.

**PROOF.** When $P,Q \in S_C$,

$$d_{\mathrm{RNTV}}(P,Q) = \frac{\gamma\left(\sum\limits_{i=1}^{n}|p_i - q_i|\right)}{\gamma\sum\limits_{i=1}^{n}\max\{p_i,q_i\} + (1-\gamma)C}.$$

Noting

$$\sum_{i=1}^{n}|p_i - q_i| \leq C \leq \sum_{i=1}^{n}\max\{p_i,q_i\}, \tag{15}$$

we obtain the inequality $d_{\mathrm{RNTV}}(P,Q) \leq \gamma$.

$d_{\mathrm{RNTV}}(P,Q) = \gamma$ holds if and only if both equalities hold in (15). □

Those corollaries above show the geometric meaning of the parameter $\gamma$.

**THEOREM 4.**

$$d_{\mathrm{RNTV}}(\lambda P, \lambda Q) = d_{\mathrm{RNTV}}(P,Q).$$

**PROOF.**

$$\begin{aligned}
&d_{\mathrm{RNTV}}(\lambda P, \lambda Q)\\
=&\frac{2\gamma(\sum\limits_{i=1}^{n}|\lambda p_i - \lambda q_i|)}{2\gamma\sum\limits_{i=1}^{n}\max\{\lambda p_i,\lambda q_i\} + (1-\gamma)\sum\limits_{i=1}^{n}(\lambda p_i+\lambda q_i)}\\
=&\frac{2\gamma(\sum\limits_{i=1}^{n}|p_i - q_i|)}{2\gamma\sum\limits_{i=1}^{n}\max\{p_i,q_i\} + (1-\gamma)\sum\limits_{i=1}^{n}(p_i+q_i)}\\
=&d_{\mathrm{RNTV}}(P,Q)
\end{aligned}$$

□

**THEOREM 5.** If $P,Q \in S_C \cup \{\mathbf{0}\}$ are determined and $P \neq Q$, then $d_{\mathrm{RNTV}}$ is increasing function for $\gamma$.

**PROOF.** let

$$g(\gamma) = \frac{2\gamma(\sum\limits_{i=1}^{n}|p_i - q_i|)}{2\gamma\sum\limits_{i=1}^{n}\max\{p_i,q_i\} + (1-\gamma)\sum\limits_{i=1}^{n}(p_i+q_i)},$$

then

$$g'(\gamma) = \frac{2(\sum\limits_{i=1}^{n}|p_i - q_i|)[\sum\limits_{i=1}^{n}(p_i+q_i)]}{\left[2\gamma\sum\limits_{i=1}^{n}\max\{p_i,q_i\} + (1-\gamma)\sum\limits_{i=1}^{n}(p_i+q_i)\right]^2} > 0.$$

□

**THEOREM 6.** If $P,Q \in S_C$, then

$$d_{\mathrm{RNTV}}(P,Q) \leq \frac{\gamma}{C}\left\|P-Q\right\|_1$$

with $\left\|X\right\|_1 := \sum\limits_{i=1}^{n}|x_i|$ for all $X \in S_C$ as usual.

**PROOF.**

$$\begin{aligned}
d_{\mathrm{RNTV}}(P,Q) =&\frac{2\gamma(\sum\limits_{i=1}^{n}|p_i - q_i|)}{2\gamma\sum\limits_{i=1}^{n}\max\{p_i,q_i\} + 2(1-\gamma)C}\\
\leq&\frac{2\gamma(\sum\limits_{i=1}^{n}|p_i - q_i|)}{2\gamma\left(\max\{\sum\limits_{i=1}^{n}p_i,\sum\limits_{i=1}^{n}q_i\}\right) + 2(1-\gamma)C}\\
=&\frac{2\gamma(\sum\limits_{i=1}^{n}|p_i - q_i|)}{2\gamma C + 2(1-\gamma)C}\\
=&\frac{\gamma}{C}\left\|P-Q\right\|_1
\end{aligned}$$

□

From the theorem above we can find the value depends on the parameters $\gamma$ and $C$ under $l_1$-norm.

## IV. Comparison between $d_{\mathrm{RNTV}}$ and $d_{\mathrm{NTV}}$

As [3] mentioned, we consider the RNA alphabet {U, C, A, G }. Code U as $(1,0,0,0)$: 1 shows the first letter U is present, 0 shows the second letter C does not appear, 0 shows the third letter A does not appear, 0 shows the fourth letter G does not appear. Thereby, C is represented as $(0,1,0,0)$, A is represented as $(0,0,1,0)$, G is represented as $(0,0,0,1)$. As a result, any codon can correspond to a fuzzy set as a point in the 12-dimensional fuzzy polynucleotide space $I^{12}$. For example, the codon CGU would be written as

$$(0,1,0,0,0,0,0,1,1,0,0,0) \in I^{12}.$$

However, there exists some cases in which no sufficient knowledge about the chemical structure of a particular sequence. One therefore may deal with base sequences not necessarily at a corner of the hypercube and some components of the fuzzy set are not either 0 or 1. For example,

$$(0.3, 0.4, 0.1, 0.2, 0, 1, 0, 0, 0, 0, 0, 1) \in I^{12}$$

expresses a codon XCG. In the case, the first letter X is unknown and corresponds to U to extent 0.3, C to the extent 0.4, A to extent 0.1, G to extent 0.2. Generally, we understand the value as the probability of X belonging to the specific RNA letter. In this opinion, the sum of the components in the point equals to 3.

(1) For the metric $d_{\mathrm{NTV}}$ (this case corresponds to $\gamma = 1$)

$$d_{\mathrm{NTV}}(\text{histidine, proline}) = d_{\mathrm{NTV}}(\text{CAU, CCG}) = 0.8$$
$$d_{\mathrm{NTV}}(\text{histidine, serine}) = d_{\mathrm{NTV}}(\text{CAU, UCG}) = 1$$
$$d_{\mathrm{NTV}}(\text{histidine, arginine}) = d_{\mathrm{NTV}}(\text{CAU, CGU}) = 0.5$$

(2) For the metric $d_{RNTV}$ with $\gamma = 0.5$

$$d_{\mathrm{RNTV}}(\text{histidine, proline}) = d_{\mathrm{RNTV}}(\text{CAU, CCG}) = 0.5$$
$$d_{\mathrm{RNTV}}(\text{histidine, serine}) = d_{\mathrm{RNTV}}(\text{CAU, UCG}) = 0.6667$$
$$d_{\mathrm{RNTV}}(\text{histidine, arginine}) = d_{\mathrm{RNTV}}(\text{CAU, CGU}) = 0.2857$$

(3) For the metric $d_{RNTV}$ with $\gamma = 0.3$

$$d_{\mathrm{RNTV}}(\text{histidine, proline}) = d_{\mathrm{RNTV}}(\text{CAU, CCG}) = 0.3333$$
$$d_{\mathrm{RNTV}}(\text{histidine, serine}) = d_{\mathrm{RNTV}}(\text{CAU, UCG}) = 0.4615$$
$$d_{\mathrm{RNTV}}(\text{histidine, arginine}) = d_{\mathrm{RNTV}}(\text{CAU, CGU}) = 0.1818$$

From the above, the $d_{\mathrm{RNTV}}$ is decreasing when the parameter $\gamma$ is decreasing. And the decreasing trend is not linear. But the value of $\gamma$ does not change the relationship of the distances between different codons. Next, the distances between codon XCG mentioned and proline and serine are

(1) For the metric $d_{NTV}$ (this case corresponds to $\gamma = 1$)

$$d_{\mathrm{NTV}}(\text{XCG, proline}) = d_{\mathrm{NTV}}(\text{XCG, CCG}) = 0.3333$$
$$d_{\mathrm{NTV}}(\text{XCG, serine}) = d_{\mathrm{NTV}}(\text{XCG, UCG}) = 0.3784$$

(2) For the metric $d_{RNTV}$ with $\gamma = 0.5$

$$d_{\mathrm{RNTV}}(\text{XCG, proline}) = d_{\mathrm{NTV}}(\text{XCG, CCG}) = 0.1818$$
$$d_{\mathrm{RNTV}}(\text{XCG, serine}) = d_{\mathrm{NTV}}(\text{XCG, UCG}) = 0.2090$$

(3) For the metric $d_{RNTV}$ with $\gamma = 0.3$

$$d_{\mathrm{RNTV}}(\text{XCG, proline}) = d_{\mathrm{NTV}}(\text{XCG, CCG}) = 0.1132$$
$$d_{\mathrm{RNTV}}(\text{XCG, serine}) = d_{\mathrm{NTV}}(\text{XCG, UCG}) = 0.1308$$

We apply the comparison to complete genomes. In[14], Torres and Nieto computed the frequencies of the nucleotides A, C, G and T at the three base sites of a codon in two bacteria *M. tuberculosis* and *E. coli*, and obtain two points corresponding to either:

$$(0.1632, 0.3089, 0.1724, 0.3556, 0.2036, 0.3145, 0.1763,$$
$$0.3056, 0.1645, 0.3461, 0.1593, 0.3302) \in I^{12}$$
$$(0.1605, 0.2420, 0.2600, 0.3374, 0.3116, 0.2286, 0.2846,$$
$$0.1752, 0.2619, 0.2568, 0.1831, 0.2981) \in I^{12}$$

For the metric $d_{NTV}$ (this case corresponds to $\gamma = 1$)

$$d_{\mathrm{NTV}}(\text{M. tuberculosis, E. coli}) = \frac{1.7012}{6.8506} = 0.2483$$

For the metric $d_{RNTV}$ with $\gamma = 0.5$

$$d_{\mathrm{RNTV}}(\text{M. tuberculosis, E. coli}) = \frac{0.8506}{6.4253} = 0.1324$$

For the metric $d_{RNTV}$ with $\gamma = 0.3$

$$d_{\mathrm{RNTV}}(\text{M. tuberculosis, E. coli}) = \frac{0.5104}{6.2552} = 0.08159$$

As a result, the various value of $\gamma$ may adjust the difference degree between two codons or complete genomes to meet our demand.

## V. Concluding Remarks

In our work, we obtain the discrimination of fuzzy sets related to the NTV metric. We mainly discuss the metric properties of the discrimination by the parameter $\gamma$ and thus a question in section 2 are settled. Thereby, we introduce a new metric $d_{\mathrm{RNTV}}$. In the future, the geometric properties of $d_{\mathrm{RNTV}}$ such as fixed point theory([15]),the isometric property and the isomorphism property are worthy to be studied. Applications of the discrimination in mathematical biology, especially in sequence analysis are also meaningful as the value of parameter $\gamma$ maybe corresponds to some biological significance.

## Acknowledgment

The authors would like to thank the editor and referees for their helpful suggestions and comments on the manuscript.

## References

[1] V. Popov, "The Approximate Period Problem," *IAENG International Journal of Computer Science*, vol. 36, no. 4, pp. 268-274, 2009.
[2] A. Gorbenko, V. Popov, "The c-Fragment Longest Arc-Preserving Common Subsequence Problem," *IAENG International Journal of Computer Science*, vol. 39, no. 3, pp. 231-238, 2012.
[3] J. J. Nieto, A. Torres and M. M. Vázquez-Trasande, "A metric space to study differences between polynucleotides," *Applied Mathematics Letters*, vol. 16, no. 8, pp. 1289-1294, 2003.
[4] B. Kosko, *Neural Networks and Fuzzy Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1992.
[5] C. T. Lin, "Adaptive subsethood for radial basis fuzzy systems." In: B. Kosko(Eds.), *Fuzzy Engineering*, Prentice-Hall, Upper Saddle River, NJ, 1997, pp. 429-464[Chapter 13].
[6] K. Sadegh-Zadeh, "Fuzzy genomes," *Artificial Intelligence in Medicine*, vol. 18, no. 1, pp. 1-28, 2000.
[7] J. J. Niteo, A. Torres, D. N. Georgiou and T. E. Karakasidis, "Fuzzy polynucleotide space and metrics," *Bulletin of Mathematical Biology*, vol. 68, no. 3, pp. 703-725, 2006.
[8] D. N. Georgiou, T. E. Karakasidis, J. J. Niteo, A. Torres, "A study of entropy/clarity of genetic sequences using metric spaces and fuzzy sets," *Journal of Theoretical Biology*, vol. 267, no. 1, pp. 95-105, 2010.
[9] A. Dress and T. Lokot, "A simple proof of the triangle inequality for the NTV metric," *Applied Mathematics Letters*, vol. 16, no. 6, pp. 809-813, 2003.
[10] B. M. Werners," Aggregation models in mathematical programming." In: G. Mitra(Eds.), *Mathematical Models for Decision Support*, NATO ASI Series, Vol. F48, Springer-Verlag, Berlin Heidelberg, 1988, pp. 295-305.
[11] B. Fustier, "The mathematical framework of fuzzy logic." In: A. Lemmi, G. Betti(Eds.), *Fuzzy Set Approach to Multidimensional Poverty Measurement*, Springer Science + Business Media, New York, 2006, pp. 29-47.
[12] M. Setnes, R. Babuška, U. Kaymak, H. R. van Nauta Lemke, "Similarity measures in fuzzy rule base simplification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 28, no. 3, pp. 376-386, 1998.
[13] F. Faghihi, "Generalization of the dissimilarity measure of fuzzy sets," *International Mathematical Forum*, vol. 2, no. 68, pp. 3395-3400, 2007.
[14] A. Torres, J. Nieto, "The fuzzy polynucleotide space: basic properties," *Bioinformatics*, vol. 19, no. 5, pp. 587-592, 2003.
[15] M. Arshad, A. Azam, P. Vetro, "Common fixed point of generalized contractive type mappings in cone metric spaces," *IAENG International Journal of Applied Mathematics*, vol. 41, no. 3, pp. 246-251, 2011.