arXiv:2302.13433v1 [math.MG] 26 Feb 2023

关于子集度量

RICHARD CASTRO、ZHIBIN CHANG、ETHAN HA、EVAN HALL 和 HIREN MAHARAJ

Abstract. 对于一个有界度量空间 X，我们在其所有有限子集构成的集合上定义一种度量。这推广了 Wentu Song、Kui Cai 和 Kees A. Schouhamer Immink [7] 提出的序列-子集距离，用于研究基于DNA的数据存储的纠错码。本文也补充了 Eiter 和 Mannila [3] 的研究；他们探讨了在各种应用背景下将距离函数扩展到空间子集上的方法。

1. 引言 Introduction

为设计适用于DNA存储信道的纠错码，文献 [7]提出了一种称为序列-子集距离的新度量。该度量将汉明距离推广为定义在任意两个无序向量集合之间的距离函数。其定义如下。设 $\mathbb{A}$ 为一个固定的有限字母表，$L \geq 1$ 为一个整数。对于任意 x_1、$x_2 \in \mathbb{A}^L$，d_H（x_1，x_2）之间的汉明距离是 x_1 与 x_2 中不同坐标的数量。对于两个子集 X_1、$X_2 \subset \mathbb{A}^L$，若 $|X_1| \leq |X_2|$，且对于任意单射 $\chi:\ X_1 \to X_2$，X_1 与 X_2 之间的 $\chi-$ 距离定义为

$$d_\chi(X_1, X_2) = \sum_{x \in X_1} d_H(x, \chi(x)) + L(|X_2| - |X_1|). \tag{1}$$

X_1 与 X_2 之间的序列-子集距离定义为

$$d_S(X_1, X_2) = d_S(X_2, X_1) = \min\{d_\chi(X_1, X_2) | \chi : X_1 \to X_2 \text{ is an injection}\}.$$

文献 [7]证明，d_S 实际上是 $\mathbb{A}^L$的子集集合上的一个度量。

本文将序列-子集距离推广如下。设 X 是一个有界度量空间。对于每个 $y \in X$，令 $M : X \to \mathbb{R}$ 为一个函数，使得

$$d(x, y) \leq M(x) \leq d(x, z) + M(z) \tag{2}$$

对所有 x、y、$z \in X$均成立。令 $Y := \mathcal{F}(X)$表示 X的所有有限子集构成的集合。对于 A、$B \in Y$，其中$|A| \leq |B|$，以及任意单射 χ: $A \to B$，A 与 B 之间的 $\chi-$ 距离定义为

$$d_\chi(A, B) := \sum_{x \in A} d(x, \chi(x)) + \sum_{y \in B \setminus \chi(A)} M(y).$$

现在，A与 B 之间的距离定义为

$$d_S(A, B) = d_S(B, A) := \min\{d_\chi(A, B) |\ \chi : A \to B \text{ is an injection}\}. \tag{3}$$

我们将在第2节证明，d_S 确实是 $\mathcal{F}(X)$上的一个度量。我们将此距离函数简称为子集度量。

在选择函数 M时有一定的灵活性。由于 X 是一个有界度量空间，我们可以选取函数 M，使其为常值函数，取值为 $D := \sup\{d(x, y) : x, y \in X\}$。对于 $X = \mathbb{A}^L$上的汉明度量 $d = d_H$，这相当于对所有 $y \in X$，令 $M(y)$ 为常数 L，由此便可得到 [7] 的子集序列度量。事实上

M 可以是任意常值函数，只要其取值是 X上的度量 d 的一个上界即可。或者，也可以如下定义 M：对于每个 $x \in X$，令

$$M(x) = \sup\{d(x, y) : y \in X\}. \tag{4}$$

条件（2）得到满足：对于所有 $y \in X$, $d(x, y) \leq d(x, z) + d(z, y) \leq d(x, z) + M(z)$，因此$M(x) \leq d(x, z) + M(z)$。

至于 [7], 的序列-子集距离， A 与 B 之间的子集距离可由一个二分图的最小权完美匹配计算得出，该二分图的两个部集分别为 A 和 B；连接 $a \in A$ 与 $b \in B$ 的边被赋予权重 $d(a, b)$。匈牙利算法可在时间 $O(|B|^3)$ [4] 内完成此操作。

该广义度量可能具有更多应用。例如，令 X 为一个有限连通图的顶点集，并令 $d(x, y)$ 为 x 与 y 之间最短路径的长度。那么， d_S 是幂集 2^X 上的一个度量，并为顶点集合之间的距离提供了一种度量方式。

另一个例子是图像识别。在这种情况下，取 X 为标准欧几里得平面中的一个有界子集（例如，对应于由像素组成的光栅）。为简单起见，我们以 $X = [0, 1] \times [0, 1]$ 单位正方形为例，并令 $d(p, q) = ||p - q||$ 为标准欧几里得距离。 X 的每个有限子集都对应一幅图像。利用 (4) 定义函数 $M(p)$，我们有 $M(p) := \max\{||p - c_1||, ||p - c_2||, ||p - c_3||, ||p - c_4||\}$ 其中c_1、 c_2、 c_3、 c_4 是 X的四个角。或者，也可以将 M 替换为值为 $D = \sqrt{2}$的常函数。

度量空间中子集之间的距离函数以及测度空间已得到广泛研究，关于这类距离的综述见 [1] ；另见 [2]。最广泛使用的子集度量之一是豪斯多夫度量 [1]。这种度量有许多变体，但我们陈述其中一个版本以作比较。设 X 为一个度量为 d的有界度量空间。对于 X的非空紧子集 A、 B ，定义

$$h(A, B) := \max\{\max_{a \in A} d(a, B), \max_{b \in B} d(b, A)\}$$

其中 $d(a, B) := \min_{a \in A} d(x, a)$， $d(b, A)$ 的定义也类似。函数 h 在 X 的所有紧子集上定义了一个度量，并将 d推广为：对所有 $a, b \in X$， $h(\{a\}, \{b\}) = d(a, b)$。如果 X 有限，豪斯多夫度量可在多项式时间内计算，并且确有理论上的优点；例如，如果 X 关于 d 完备，那么它也是完备的。不过，正如 [3], 所指出的，它可能不适用于某些应用，因为该度量没有考虑某些有限集合的整体构型。另一方面， [7] 为比较 DNA序列而提出的子集序列度量，能够更精细地比较两组序列，因此在这种情况下是更合适的距离度量。式（1）右侧每个涉及 L 的项，都表示一条 DNA链与另一集合相距过远时所对应的自然最坏情况权重。尽管这项工作的作者主要是受到将 [7], 的工作加以推广的动机所驱使，这项工作也补充了 [3] 的研究；后者探讨了如何更一般地将距离度量扩展到子集。作为比较，我们简要回顾 [3] 的一些主要结果。非空集合 B 上的距离函数 Δ 满足度量的所有公理，但可能不满足三角不等式。在 [3], 中，作者研究了如何将距离函数扩展到 B 的非空有限子集集合上的问题。他们还讨论了计算此类扩展的算法。为了度量两个

B 的非空子集 S_1, S_2 之间的距离，他们讨论了四种距离函数：最小距离之和 [5]

$$d_{md}(S_1, S_2) := \frac{1}{2}\left(\sum_{e\in S_1}\Delta(e, S_2) + \sum_{e\in S_2}\Delta(e, S_1)\right),$$

满射距离

$$d_s(S_1, S_2) := \min_{\eta}\sum_{(e_1,e_2)\in\eta}\Delta(e_1, e_2)$$

其中，最小值是在从较大集合到较小集合的所有满射 η 中取的（归功于 G. Oddie 在 [6]中的工作），即公平满射距离

$$d_{fs}(S_1, S_2) := \min_{\eta}\sum_{(e_1,e_2)\in\eta}\Delta(e_1, e_2)$$

其中，最小值是在从较大集合到较小集合的所有公平满射 η 中取的（若满射 η：$S_1 \to S_2$ 满足对于所有 $x, y \in S_1$，$||\eta^{-1}(x)| - |\eta^{-1}(y)|| \leq 1$，则称其为公平满射；这一概念同样归功于 G. Oddie 在 [6]中的工作），并且他们还引入了链接距离

$$d_l(S_1, S_2) := \min_{R}\sum_{(e_1,e_2)\in R}\Delta(e_1, e_2)$$

其中，最小值是在 S_1 与 S_2 之间的所有链接关系 R 中取的（若子集 $R \subset S_1 \times S_2$满足对于所有 $e_1 \in S_1$，存在 $e_2 \in S_2$ 使得 $(e_1, e_2) \in R$，并且对于所有 $e_2 \in S_2$，存在 $e_1 \in S_1$ 使得 $(e_1, e_2) \in R$，则称其为链接关系）。虽然他们表明，当 B 是整数平面的有限子集且 Δ 是曼哈顿度量时，这些距离函数不满足度量的定义，但 Eiter 和 Mannila 提出了一种优雅的构造，称为度量下确界方法，它可以根据给定的距离函数Δ构造出一个度量 Δ^ω。有趣的是，他们证明了 $d_s^\omega = d_{fs}^\omega = d_l^\omega$。 [3] 中的作者认为，在某些情境下，链接度量非常直观。研究这种度量在 DNA 数据存储纠错码中的应用也会很有意思。

本文其余部分将致力于证明 (3) 确实是一个度量。

2. Proofs

尽管本节中 X 是一个度量为 d的有界度量空间，函数 M：$X \to \mathbb{R}$ 满足条件 (2)，d_S 是由 (3) 定义的函数，而 $\mathcal{F}(X)$ 是 X的所有有限子集构成的集合。本节将证明函数 d_S 是 $\mathcal{F}(X)$ 上的一个度量。虽然本文所遵循的主要步骤受 [7], 的启发，但由于函数 M 出现在 d_S的定义中，仍需作出相应调整。

引理 1. 对于任意 $X_1, X_2 \in \mathcal{F}(X)$，若满足 $|X_1| \leq |X_2|$，则存在一个单射 χ_0:$X_1 \to X_2$，使得对于所有 $x \in X_1 \cap X_2$，$d_S(X_1, X_2) = d_{\chi_0}(X_1, X_2)$ 且 $\chi_0(x) = x$。

证明。若 $X_1 \cap X_2 = \emptyset$，则该命题显然成立。设 $X_1 \cap X_2 \neq \emptyset$。取 χ： $X_1 \to X_2$，使得 $d_S(X_1, X_2) = d_\chi(X_1, X_2)$。证明分两部分。首先证明，必要时可在 $X_1 \cap X_2$ 上重新定义 χ，使得 $d_S(X_1, X_2) = d_\chi(X_1, X_2)$，且 $X_1 \cap X_2$ 包含在 χ 的像中。接着证明，可进一步调整 χ，使其具有所需的性质。

设某个 $x_0 \in X_1 \cap X_2$ 不属于 χ 的像。于是，我们在 x_0 处重新定义 χ，构造一个新的嵌入 $\nu: X_1 \to X_2$，定义如下:

$$\nu(x) = \begin{cases} \chi(x) & \text{if } x \neq x_0 \\ x_0 & \text{if } x = x_0. \end{cases}$$

根据定义 $d_S(X_1, X_2) \leq d_\nu(X_1, X_2)$。注意，$\nu(X_1) = (\chi(X_1) \setminus \{\chi(x_0)\}) \cup \{x_0\}$ 以及

$$\sum_{x \in X_1} d(x, \nu(x)) = \sum_{x \in X_1} d(x, \chi(x)) - d(x_0, \chi(x_0)). \tag{5}$$

由于 $x_0 \in X_2 \setminus \chi(X_1)$，$\chi(x_0) \notin X_2 \setminus \chi(X_1)$ 且 $\chi(x_0) \in X_2 \setminus \nu(X_1)$，因此

$$\sum_{y \in X_2 \setminus \nu(X_1)} M(y) = \sum_{y \in X_2 \setminus \chi(X_1)} M(y) - M(x_0) + M(\chi(x_0)). \tag{6}$$

结合（5）和（6），我们得到

$$d_\nu(X_1, X_2) = d_\chi(X_1, X_2) + M(\chi(x_0)) - M(x_0) - d(x_0, \chi(x_0)).$$

由条件（2）可知，$d_\nu(X_1, X_2) \leq d_\chi(X_1, X_2) = d_S(X_1, X_2)$。因此，$d_S(X_1, X_2) = d_\nu(X_1, X_2)$ 且 $\nu(x_0) = x_0$。重复应用上述过程，我们将得到一个从 X_1 到 X_2 的嵌入，我们也将其称为 χ，并具有如下性质：$X_1 \cap X_2 \subseteq Im(\chi)$。

设 $x_1 \in X_1 \cap X_2$。接下来我们证明，如果 $\chi(x_1) \neq x_1$，那么可以调整嵌入χ，构造一个新的嵌入 $\mu: X_1 \to X_2$，使得 $\mu(x_1) = x_1$，并且仍有$d_S(X_1, X_2) = d_\mu(X_1, X_2)$。由上文可知，存在 $z \in X_1$ 使得 $\chi(z) = x_1$。令 $y = \chi(x_1)$，并定义

$$\mu(x) = \begin{cases} \chi(x) & \text{if } x \neq x_1, z \\ x_1 & \text{if } x = x_1 \\ y & \text{if } x = z. \end{cases}$$

于是 $\mu: X_1 \to X_2$ 是一个单射，并且根据子集距离的定义，$d_S(X_1, X_2) \leq d_\mu(X_1, X_2)$。此外，我们有

$$\begin{aligned} d_\chi(X_1, X_2) &= d(x_1, y) + d(z, x_1) + (d_\mu(X_1, X_2) - d(x_1, x_1) - d(z, y)) \\ &= d_\mu(X_1, X_2) + d(x_1, y) + d(z, x_1) - d(z, y) \\ &\geq d_\mu(X_1, X_2) \end{aligned}$$

其中，最后一个不等式由三角不等式得出。因此 $d_S(X_1, X_2) \geq d_\mu(X_1, X_2)$，并且我们看到 $d_S(X_1, X_2) = d_\chi(X_1, X_2) = d_\mu(X_1, X_2)$，且 $\mu(x_1) = x_1$。重复应用上述过程，我们便得到一个具有所需性质的嵌入。□

推论 1. 对于任意 $X_1, X_2 \in \mathcal{F}(X)$，

$$d_S(X_1, X_2) = d_S(X_1 \setminus X_2, X_2 \setminus X_1).$$

*Pr*证明。这是引理 1 和 $d_\chi(\cdot, \cdot)$ 定义的直接推论 . □

引理2. 假设 $X_1, X_2 \in \mathcal{F}(X)$，其中 $|X_1| \leq |X_2|$。那么对于任意 $b \in X$, $d_S(X_1, X_2)$，有 $\leq d_S(X_1, X_2 \cup \{b\})$。

证明. 假设 $\chi\colon X_1 \to X_2 \cup \{b\}$，使得 $d_S(X_1, X_2 \cup \{b\}) = d_\chi(X_1, X_2 \cup \{b\})$。若$\chi(X_1) \subseteq X_2$，则 $d_\chi(X_1, X_2 \cup \{b\}) = d_\chi(X_1, X_2) + M(b) \geq d_S(X_1, X_2) + M(b) \geq d_S(X_1, X_2)$。若 $\chi(X_1) \not\subset X_2$，则 $\chi(a) = b$，其中某个 $a \in X_1$ 和 $|X_2| > |X_1|$。固定 $c \in X_2 \setminus \chi(X_1)$，并定义 $\eta\colon X_1 \to X_2$ 为

$$\eta(x) = \begin{cases} \chi(x) & \text{if } x \neq a \\ c & \text{if } x = a. \end{cases}$$

于是 $\eta(X_1) = (\chi(X_1) \setminus \{b\}) \cup \{c\}$ 因此 $X_2 \cup \{b\} \setminus \chi(X_1)$ 是不相交并集 $(X_2 \setminus \eta(X_1)) \cup \{c\}$ 且

$$\begin{aligned}
& d_S(X_1, X_2 \cup \{b\}) \\
= \; & d_\chi(X_1, X_2 \cup \{b\}) \\
= \; & \sum_{x \in X_1} d(x, \chi(x)) + \sum_{y \in X_2 \cup \{b\} \setminus \chi(X_1)} M(y) \\
= \; & d(a, b) + \sum_{x \in X_1} d(x, \eta(x)) - d(a, c) + \sum_{y \in X_2 \setminus \eta(X_1)} M(y) + M(c) \\
= \; & d(a, b) + M(c) - d(a, c) + \sum_{x \in X_1} d(x, \eta(x)) + \sum_{y \in X_2 \setminus \eta(X_1)} M(y) \\
= \; & d(a, b) + M(c) - d(a, c) + d_\eta(X_1, X_2) \\
\geq \; & d_\eta(X_1, X_2) \geq d_S(X_1, X_2)
\end{aligned}$$

由于由条件 (2) 可知 $d(a, c) \leq M(c)$。 □

By 反复应用上述结果，我们得到以下推论 y.

推论 2. 对任意 $X_1, X_2 \in \mathcal{F}(X)$，满足 $|X_1| \leq |X_2|$。假设 $X_2' \subseteq X_2$ 满足 $|X_1| \leq |X_2'|$。那么

$$d_S(X_1, X_2') \leq d_S(X_1, X_2).$$

定理 1. $d_S(\cdot, \cdot)$ 是 $\mathcal{F}(X)$ 上的一个度量。

证明。对于两个有限集 A 和 B，我们用 $\mathscr{X}(A, B)$ 表示单射 $\chi\colon A \to B$ 的集合。设 $X_1, X_2 \in \mathcal{F}(X)$。根据 $d_S(\cdot, \cdot)$ 的定义，我们有 $d_S(X_1, X_2) = d_S(X_2, X_1) \geq 0$。我们证明 $d_S(X_1, X_2) = 0$ 当且仅当 $X_1 = X_2$。不妨设 $|X_1| \leq |X_2|$，并令 $\nu \in \mathscr{X}(X_1, X_2)$ 满足 $d_S(X_1, X_2) = d_\nu(X_1, X_2)$。那么 $d_S(X_1, X_2) = d_\nu(X_1, X_2) = 0$ 当且仅当 $\sum_{x \in X_1} d(x, \nu(x)) + \sum_{y \in X_2 \setminus \nu(X_1)} M(y) = 0$ 当且仅当 $d(x, \nu(x)) = 0$ 对所有 $x \in X_1$ 都成立，且 $X_2 = \nu(X_1)$ 当且仅当 $x \in X_1$

对于所有 $x \in X_1$ 和 $|X_2| = |X_1|$，$x = \nu(x)$ 当且仅当 $X_1 = X_2$。

因此，我们只需证明 $d_S(\cdot, \cdot)$ 满足三角不等式。设 $X_1, X_2, X_3 \in \mathcal{F}(X)$。我们将通过考察各种情况证明 $d_S(X_1, X_2) \leq d_S(X_1, X_3) + d_S(X_3, X_2)$。注意，我们仍假设 $|X_1| \leq |X_2|$，并且 $\nu \in \mathscr{X}(X_1, X_2)$ 满足 $d_S(X_1, X_2) = d_\nu(X_1, X_2)$。

情形 1：假设 $|X_1| \leq |X_3| \leq |X_2|$。令 $\mu \in \mathscr{X}(X_3, X_2)$ 和 $\eta \in \mathscr{X}(X_1, X_3)$ 满足 $d_S(X_3, X_2) = d_\mu(X_3, X_2)$，且 $d_S(X_1, X_3) = d_\eta(X_1, X_3)$。我们可以假设

$$\begin{aligned}
X_1 &= \{x_1, \ldots, x_n\} \\
X_3 &= \{y_1, \ldots, y_n, y_{n+1}, \ldots, y_{n+s}\} \\
X_2 &= \{z_1, \ldots, z_n, z_{n+1}, \ldots, z_{n+s}, \ldots, z_{n+s+t}\}
\end{aligned}$$

其中 s、$t \geq 0$ 和 $\mu(y_i) = z_i$ 对 $1 \leq i \leq n+s$ 成立，而 $\eta(x_i) = z_i$ 对 $1 \leq i \leq n$ 成立。于是

$$d_S(X_1, X_3) = \sum_{i=1}^{n} d(x_i, y_i) + \sum_{i=n+1}^{n+s} M(y_i) \text{ and}$$
$$d_S(X_2, X_3) = \sum_{i=1}^{n+s} d(y_i, z_i) + \sum_{i=n+s+1}^{n+s+t} M(z_i).$$

设 $\chi = \mu \circ \eta \in \mathscr{X}(X_1, X_2)$。则

$$\begin{aligned}
& d_S(X_1, X_2) \\
\leq & d_\chi(X_1, X_2) \\
= & \sum_{i=1}^{n} d(x_i, z_i) + \sum_{i=n+1}^{n+s+t} M(z_i) \\
\leq & \sum_{i=1}^{n} [d(x_i, y_i) + d(y_i, z_i)] + \sum_{i=n+1}^{n+s+t} M(z_i) \\
= & \sum_{i=1}^{n} d(x_i, y_i) + \sum_{i=1}^{n} d(y_i, z_i) + \sum_{i=n+1}^{n+s+t} M(z_i) \\
= & \left(d_S(X_1, X_3) - \sum_{i=n+1}^{n+s} M(y_i) \right) + \\
& \left(d_S(X_3, X_2) - \sum_{i=n+1}^{n+s} d(y_i, z_i) - \sum_{i=n+s+1}^{n+s+t} M(z_i) \right) + \sum_{i=n+1}^{n+s+t} M(z_i) \\
= & d_S(X_1, X_3) + d_S(X_2, X_3) - \sum_{i=n+1}^{n+s} M(y_i) - \sum_{i=n+1}^{n+s} d(y_i, z_i) + \sum_{i=n+1}^{n+s} M(z_i) \\
= & d_S(X_1, X_3) + d_S(X_2, X_3) + \sum_{i=n+1}^{n+s} (M(z_i) - M(y_i) - d(y_i, z_i)) \\
\leq & d_S(X_1, X_3) + d_S(X_3, X_2).
\end{aligned}$$

由条件 (2) 可得

情形 2：假设 $|X_3| \leq |X_1| \leq |X_2|$。令 $\mu \in \mathscr{X}(X_3, X_2)$ 和 $\eta \in \mathscr{X}(X_3, X_1)$ 满足 $d_S(X_3, X_2) = d_\mu(X_3, X_2)$ 且 $d_S(X_1, X_3) = d_\eta(X_1, X_3)$。我们可以假设

$$\begin{aligned}
X_3 &= \{x_1, \ldots, x_n\} \\
X_1 &= \{y_1, \ldots, y_n, y_{n+1}, \ldots, y_{n+s}\} \\
X_2 &= \{z_1, \ldots, z_n, z_{n+1}, \ldots, z_{n+s}, \ldots, z_{n+s+t}\}
\end{aligned}$$

其中 s、$t \geq 0$ 和 $\mu(x_i)$ 对于 $1 \leq i \leq n$ 为 $= z_i$，而 $\eta(x_i)$ 对于 $1 \leq i \leq n$ 为 $= y_i$。于是

$$d_S(X_3, X_1) = \sum_{i=1}^{n} d(x_i, y_i) + \sum_{i=n+1}^{n+s} M(y_i)$$
$$d_S(X_3, X_2) = \sum_{i=1}^{n} d(x_i, z_i) + \sum_{i=n+1}^{n+s+t} M(z_i).$$

定义 χ： $X_1 \to X_2$ 由 $\chi(y_i) = z_i$ 给出，其中 $i = 1, 2, \ldots, n+s$。那么

$$\begin{aligned}
& d_S(X_1, X_2) \\
\leq & d_\chi(X_1, X_2) \\
= & \sum_{i=1}^{n+s} d(y_i, z_i) + \sum_{i=n+s+1}^{n+s+t} M(z_i) \\
= & \sum_{i=1}^{n} d(y_i, z_i) + \sum_{i=n+1}^{n+s} d(y_i, z_i) + \sum_{i=n+s+1}^{n+s+t} M(z_i) \\
\leq & \sum_{i=1}^{n} [d(y_i, x_i) + d(x_i, z_i)] + \sum_{i=n+1}^{n+s} d(y_i, z_i) + \sum_{i=n+s+1}^{n+s+t} M(z_i) \\
= & \sum_{i=1}^{n} d(y_i, x_i) + \left(\sum_{i=1}^{n} d(x_i, z_i) + \sum_{i=n+1}^{n+s+t} M(z_i) \right) - \sum_{i=n+1}^{n+s} M(z_i) + \sum_{i=n+1}^{n+s} d(y_i, z_i) \\
= & \left(d_S(X_3, X_1) - \sum_{i=n+1}^{n+s} M(y_i) \right) + d_S(X_3, X_2) + \sum_{i=n+1}^{n+s} (d(y_i, z_i) - M(z_i)) \\
= & d_S(X_3, X_1) + d_S(X_3, X_2) - \sum_{i=n+1}^{n+s} M(y_i) + \sum_{i=n+1}^{n+s} (d(y_i, z_i) - M(z_i)) \\
\leq & d_S(X_1, X_3) + d_S(X_3, X_2).
\end{aligned}$$

最后一个不等式由条件 (2) 得出。

情形 3：假设 $|X_1| \leq |X_2| \leq |X_3|$。

取 X_3 的一个基数为 X_2 的子集 X_3'。由情形 1 可知，$d_S(X_1, X_2) \leq d_S(X_1, X_3') + d_S(X_3', X_2)$。由推论 2 可知，$d_S(X_1, X_3') \leq d_S(X_1, X_3)$，且 $d_S(X_3', X_2) \leq d_S(X_3, X_2)$。因此，$d_S(X_1, X_2) \leq d_S(X_1, X_3) + d_S(X_3, X_2)$。

□

注 1. 如果 X 至少包含两个元素，那么函数 M 永远不会取值 0。事实上，存在一个常数 $C > 0$，使得对所有 $y \in X$，$M(y) \geq C$：由 (2) 可得，$d(x, y) \leq M(x) \leq d(x, y) + M(y) \leq 2M(y)$。因此，对所有 $y \in X$，$M(y) \geq M(x)/2$。令 $C = M(x)/2$。如果 $C = 0$，那么不等式 $M(y) \geq d(y, x)$ 意味着对所有 $x \in X$，$y = x$，这与 X 至少包含两个元素相矛盾。因此，$C = M(x)/2 > 0$ 就是所需的常数。

注 2. 如果 $\{A_n\}$ 是 $\mathcal{F}(X)$ 中的柯西序列，可以证明，当 m, n 足够大时，对所有 m，n 都有 $|A_n| = |A_m|$：令 C 如注 1 所述。于是存在 N，使得7

$d_S(A_m, A_n)$ 对所有 $m, n \geq N$ 都有 $< C$。由于对所有 $y \in X$，$C = \frac{1}{2}M(x) < M(y)$，因此对所有 $m, n \geq N$，都有$|A_m| = |A_n|$。

注 3. 如果由度量 d 在 X 上诱导的拓扑是离散拓扑，那么 $\mathcal{F}(X)$ 关于子集度量是完备的。然而，一般情况下并非如此。考虑这样的情形：$X = [0, 1]$, d 是通常的欧几里得度量，且 $M(y) = \max\{y, 1 - y\}$。令$A_n = \{0, \frac{1}{n}\}$ 对所有 $n \geq 1$成立。于是 $\{A_n\}$ 是一个不收敛的柯西序列：如果 $\{A_n\}$确实收敛，那么根据引理 1 和注 2，它将收敛到某个形如 $A = \{0, a\}$的集合，其中 $a \in X$。但当 $n \to \infty$时，$d_S(A_n, A) = |a - 1/n| \to 0$，因此 a 必定等于 0。然而，如果 $a = 0$，那么当 $n \to \infty$时，$d_(A_n, A) = M(1/n) = 1 - 1/n \to 1$，矛盾。

References

[1] Aura Conci 和 Carlos Kubrusly。集合间的距离——综述。Adv. Math. Sci. Appl.，26(1):1–18，2017。[2] Michel Marie Deza 和 Elena Deza。距离百科全书。Springer-Verlag，柏林，2009。附 1 张 CD-ROM（Windows、Macintosh 和 UNIX）。[3] Thomas Eiter 和 Heikki Mannila。点集的距离度量及其计算。Acta Inform.，34(2):109–133，1997。[4] James Munkres。指派和运输问题的算法。J. Soc. Indust. Appl. Math.，5:32–38，1957。[5] Ilkka Niiniluoto。《似真性》，Synthese Library 第 185 卷。Springer Dordrecht，1987。[6] Ilkka Niiniluoto 和 Raimo Tuomela，编者。《科学变革的逻辑与认识论》。Societas Philosophica Fennica，赫尔辛基，1979。Acta Philos. Fenn. 30 (1978)，第 2-4 期（1979）。[7] Wentu Song、Kui Cai 和 Kees A. Schouhamer Immink。基于 DNA 的数据存储中用于差错控制的序列-子集距离和编码。IEEE Trans. Inform. Theory，66(10):6048–6065，2020。

Richard Castro, Department of Mathematics and Statistics,
San Diego State University,5500 Campanile Drive, San Diego, 92182, CA, USA电子邮箱: rcastro0899@sdsu.eduZhibin Chang, Department of Mathematics and Statistics,
San Diego State University,5500 Campanile Drive, San Diego, 92182, CA, USA电子邮箱: zchang@sdsu.edu

Ethan Ha, Department of Mathematics and Statistics, San Diego State University, 5500Campanile Drive, San Diego, 92182, CA, USA电子邮箱：ethankaweiha@gmail.com

Evan Hall, Department of Mathematics and Statistics, San Diego State University, 5500Campanile Drive, San Diego, 92182, CA, USA电子邮箱：elhall@sdsu.edu

Hiren Maharaj, Department of Mathematics and Statistics, San Diego State University, 5500 Campanile Drive, San Diego, 92182, CA, USA电子邮箱：hmaharaj@sdsu.edu