

絶対近傍の被覆率と点配置

大西 建輔*

星 守†

* 東海大学理学部情報数理学科

† 電気通信大学大学院情報システム学研究所

KensukeOnishi@acm.org

hoshi@is.uec.ac.jp

概要： ある母点集合の L_p 距離によるボロノイ図を考える。ある母点 q に対し、 L_1 距離から L_∞ 距離までの全てのボロノイ領域の交わりを q の絶対近傍と呼ぶ。母点集合を固定しても、用いる距離によりボロノイ領域の形状は変化し、絶対近傍の全体は必ずしも空間の分割とはならない。

本稿では、実際に絶対近傍を構築し、絶対近傍により被覆される領域の体積と空間全体の体積の比 (被覆率と呼ぶ) を実験的に評価し、絶対近傍を構築するための母点集合として、格子点を用いると、被覆率が高くなることも実験的に示した。さらに、正方格子点を用いると、絶対近傍の全体が空間の分割になることを理論的に示した。

Cover Ratio of Absolute Neighbor and Point Configuration

Kensuke Onishi*

Mamoru Hoshi†

*Department of Mathematical Science, † Graduate School of Information Systems,
Tokai University The University of Electro-Communications

KensukeOnishi@acm.org

hoshi@is.uec.ac.jp

Abstract : Fix a set of points. Voronoi diagram for the set using L_p norm is considered. For a point q in the set, let *absolute neighbor* be the intersection of Voronoi regions of q using all L_p norms. Since the shape of Voronoi region is dependent on the used norm, absolute neighbors of the set does not always cover whole space.

In this paper, we construct absolute neighbors and computed the ratio of the volume covered by all absolute neighbors to that of the whole space, called *cover ratio* by computational experiments. Computational experiments showed that the cover ration is higher when lattice is used as a set of points generating absolute neighbors than when a set of random points is used. Moreover, we show that the absolute neighbors for square lattice cover the whole space.

1 はじめに

ボロノイ図は、生物学、視覚認識、物理学、考古学等の様々な分野に応用することのできる幾何構造である。ボロノイ図の理論的側面は、計算幾何の分野で構成から性質まで非常に多数の研究がなされている。それらの研究は、Okabe らの Spatial Tessellations[4] にまとめられている。

ボロノイ図の応用の1つとして、データベースの索引の構築を挙げることができる。データの空間は、ベクトル空間として表されているとする。データベースの索引の構築時には、ベクトル空間は部分

領域に分割され、本稿では、この分割にボロノイ図の概念を用いる。そして、 ε 近傍検索を行う際、索引を用いて、データが含まれない部分領域を検索対象から削除することで、検索の高速化を行なう。索引の詳細は [1, 2] によくまとめられている。

我々は、[6] において、任意の L_p 距離での検索が可能な索引構造 mm-GNAT を提案し、検索効率についての計算機実験を行った。mm-GNAT は、検索対象の絞込みに用いる距離の範囲 (存在範囲と呼ぶ) を L_∞ 距離での最小値と L_1 距離での最大値に拡張することで、任意の L_p 距離関数に対応した。しかしながら、存在範囲の拡張を行うと、絞込みの効果

が落ち、検索効率の向上が得られない場合があった。そこで、検索効率の向上を目指し、分割に用いるボロノイ図の母点の配置について研究を行う。この点配置の満たすべき要件としては、次のことがあげられる。

- 点配置の各点から他の母点のボロノイ領域までの L_p 距離が、どのような L_p 距離を用いても、あまり変化しない。
- 存在範囲における L_∞ 距離での最小値と L_1 距離での最大値の差がなるべく小さい。

これらの条件を満たすためには、それぞれの領域が母点の周辺あり、領域内のどの点までの L_p 距離も小さい必要がある。そこで、我々は絶対近傍 (absolute neighbor)[5] を用いることを考えた。しかし、絶対近傍の全体は必ずしも空間全体を被覆しないため、一般の点配置や特殊な点配置でどの程度空間が被覆されるのかを実験的に調べた。

本稿の構成は以下の通りである。2 節では、絶対近傍の定義とそれに関する結果を述べる。3 節では、ランダムな点を母点集合としたときの絶対近傍の被覆率の実験結果を述べる。4 節では、母点を三角格子、正方格子、面心格子状に配置したときの絶対近傍についての考察を行う。これらの格子での絶対近傍の被覆率はランダムな点を母点集合とした場合よりも高いことを実験的に示す。特に、正方格子に対する計算機実験では被覆率がほぼ 1 となった。また、正方格子による絶対近傍が空間全体を被覆することを理論的に示す。

2 絶対近傍

2.1 絶対近傍の定義

絶対近傍とは、次で定義される集合である。

定義 1 (絶対近傍, [5]) $Q = \{q_1, q_2, \dots, q_n\}$ を点集合とする。このとき、点 q_i の絶対近傍を次で定義する。

$$\text{AN}(q_i; Q) := \bigcap_{p=1}^{\infty} \text{Vor}_p(q_i; Q)$$

ただし、 $\text{Vor}_p(q_i; Q)$ は、点集合 Q に対する L_p 距離

による点 q_i のボロノイ領域とする。

絶対近傍は、どの L_p 距離を用いても q_i が最近点となる点のなす領域である。絶対近傍を定義から計算するためには、無限の集合の交わりを計算しなければならない。

Onishi[5] により、次の定理が示され、絶対近傍を計算することができるようになった。

定理 1 $Q = \{q_1, q_2, \dots, q_n\}$ を点集合とする。絶対近傍 $\text{AN}(q_i; Q)$ は次で表わされる。

$$\text{AN}(q_i; Q) = \text{Vor}_1(q_i; Q) \cap \text{Vor}_\infty(q_i; Q)$$

2.2 絶対近傍による被覆

絶対近傍の全体を用いて、空間を被覆することを考える。例えば、図 1 は 2 次元での絶対近傍の一例である。この例では、10 点からなる母点集合を考え、その絶対近傍を表示している。図中の白い部分は、それぞれの母点の絶対近傍である。ここに含まれる点はこの L_p 距離を用いても、同じ母点が最近点となる。図中の黒い部分は、 L_p 距離によって最近点となる母点異なる領域である。このようにどこの絶対近傍にも含まれない部分、**非絶対近傍**と呼ぶ。本稿では、絶対近傍領域がどの程度存在するのかを計算機実験により調べる。また、絶対近傍領域ができる限り空間を被覆できるような点配置としてどのような点配置があるのかについても考察する。

3 計算機実験

本節では、計算機実験の方法、及び計算機実験の結果について述べる。絶対近傍領域がどの程度空間を被覆するかを計測するために、絶対近傍領域の体積を調べ、空間全体の体積との比を考える。一般に領域の体積を計算することは困難であるため、モンテカルロ法による近似を行なう。

まず、次元 d 、点の取りうる座標の最大値 max 、母点数 spn を決める。超立方体 $[1, max]^d$ から spn 個の母点を選び、母点集合 Q とする。超立方体に含まれる点 x に対し、 L_1 距離と L_∞ 距離が最小となる母点をそれぞれ計算する。もし、それらの母点が一致すれば、 x をその母点の絶対近傍の中の点と判定

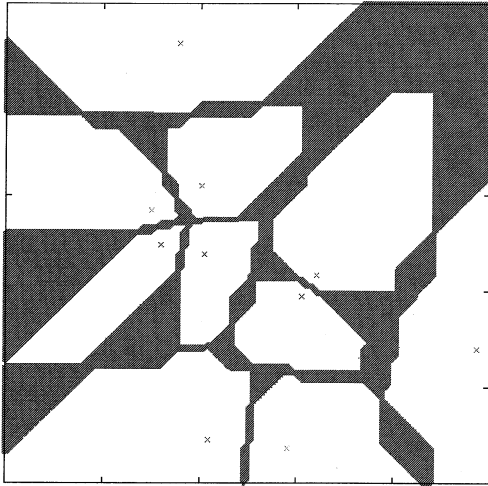


図1 絶対近傍の一例

する。この手続きが終了すると、すべての点が絶対近傍か、非絶対近傍のいずれかに分類される。図1は、2次元, $max = 1000$, $spn = 10$ の場合の絶対近傍 (白い部分) と非絶対近傍 (黒い部分) を表示している。

さて、絶対近傍がどの程度空間を被覆するのかを定量的に調べるために、**被覆率**を次で定義する。

$$\frac{\sum_i (q_i \text{の絶対近傍に含まれる点の数})}{\text{点の総数}}$$

今回の実験では、母点集合の取り方に依存しないように、10種類の母点集合を選び、被覆率の平均を計算した。その結果が表1である。このデータを片対数グラフにしたのが、図2である。図2の縦軸は被覆率、横軸は最大値の対数軸となっている。また、図3では、横軸は全体の点数に対する母点集合の数の比率となっている。

図2から、次元が2の場合では、母点数、最大値に依存せず被覆率がほぼ0.8という結果が得られている。また、4次元の場合では、ほぼ0.6、8次元の場合では、0.4前後の被覆率となっている。4次元、8次元での実験数が少ないため現在実験を進めているところである。

空間の分割として、絶対近傍の全体を用いるため

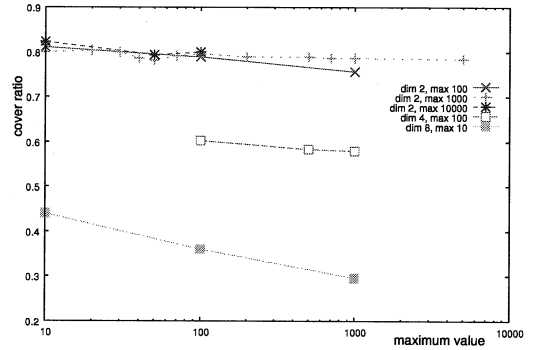


図2 被覆率 vs 座標の最大値

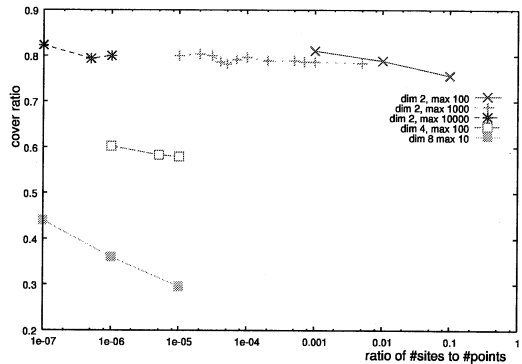


図3 被覆率 vs 全点数に対する母点数

には、空間全体を被覆する必要がある。そのための方法として、母点集合を多重化するという手法が考えられる。ある母点集合 Q の絶対近傍で、空間全体が被覆されない場合に、新たに母点集合 Q' をとり、その絶対近傍を被覆に加える。もし、空間全体が被覆されれば、空間の分割として用い、被覆されなければ、さらに母点集合 Q'' をとり、絶対近傍を追加するという手法である。以下では、被覆率 r_d ($0 \leq r_d \leq 1$) を用いて、必要となる母点集合の個数 k を見積もる。1つの母点集合で、 r_d だけ空間が被覆されるとする。被覆されるべき空間に含まれる総点数を N とすると、次の式が成立する。

$$(1 - r_d)^k \cdot N < 1.$$

表 1 被覆率, 近似値

次元	最大値	10	20	30	40	50	70	100	200	500	700	1000	5000
2	10^2	0.8109						0.7883				0.7558	
	10^3	0.7999	0.8025	0.7996	0.7867	0.7807	0.7906	0.7956	0.7890	0.7889	0.7863	0.7863	0.7839
	10^4	0.8230				0.7938		0.7990					
4	10^2							0.6025		0.5824		0.5786	
8	10^1	0.4401						0.3604				0.2962	

これを k について解くと次の評価式が得られる.

$$k > -\log_{1-r_d} N.$$

$1 - r_d$ が 1 未満の値をとるため, 右辺の対数の値は負となり, 右辺全体は正となる. 実際に, $r_d = 0.8$, $N = 10^4$ として計算を行なうと, $k > 5.7227 \dots$ という値が得られ, 6 個の母点集合をとればよいことがわかる. この評価は, 最悪の場合であり, 点集合の構成方法により, 減らすことは可能である.

4 点配置

前節では, 一般の点配置に対する被覆率の実験を行なった. 低次元の場合でも, 絶対近傍の全体によって, 必ずしも全空間が被覆されるわけではないことがわかった. そこで本節では, 被覆率が高い点配置についての考察を行なう.

点配置として, 3 種類の格子点を元にした配置について, 実験を行なった.

- 三角格子配置
- 正方格子配置
- 面心格子配置

その結果得られた絶対近傍が図 4 から図 6 である.

三角格子配置

三角格子配置とは, 点が正三角をなすように配置した点配置である. この配置に対する絶対近傍の一例を図 4 に示す. この例は, 2 次元, 最大値 100, 母点数 22 の場合であり, その被覆率は, 0.8852 となった.

正方格子配置

正方格子配置とは, 点集合を正方形形状に配置した点配置である. この配置に対する絶対近傍の一例を

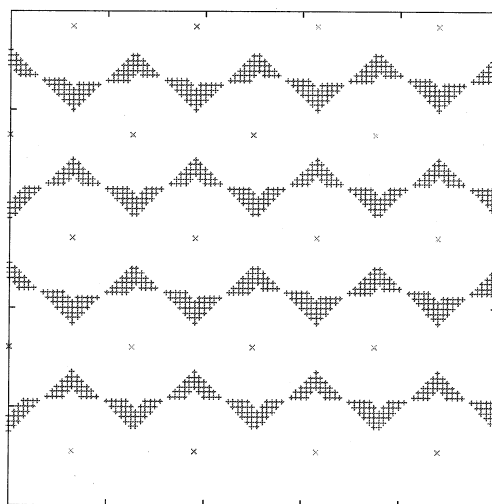


図 4 三角格子配置に対する絶対近傍

図 5 に示す. この例は, 2 次元, 最大値 100, 母点数 16 の場合であり, その被覆率は, 0.9994 となっている. また, 4 次元, 最大値 100, 点数 16 の場合では, 被覆率 1.0000 となる場合があった. 全ての実験で, 被覆率はほぼ 1 になったため, 空間のほとんどが被覆されていることがわかった.

以下では, 正方格子配置を用いた場合に, 全ての空間が絶対近傍で被覆されることを示す.

まず, 次の補題を示す.

補題 1 正方格子配置に対する L_1 距離によるボロノイ領域と L_∞ 距離によるボロノイ領域は一致する.

証明: 正方格子配置に対する L_1 距離と L_∞ 距離の垂直二等分面が一致することを示せば十分である.

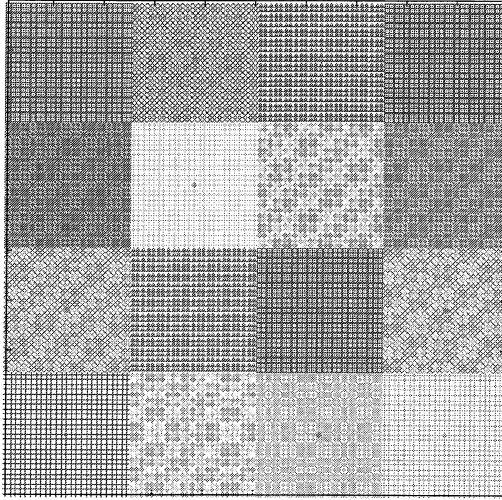


図5 正方格子配置に対する絶対近傍

今, d 次元空間の正方格子配置として,

$$\{(\overbrace{2\mathbb{Z}a, 2\mathbb{Z}a, \dots, 2\mathbb{Z}a}^d) \mid a > 0\}$$

を考える. ただし, $\mathbb{Z}$ は整数とする. 一般性を失わず, 2 格子点の間の垂直二等分面が L_1 距離と L_∞ 距離で一致することを示せばよい.

L_1 距離の場合 2 点

$$(0, 0, \dots, 0), (0, \dots, 0, 2a, 0, \dots, 0)$$

を考える. このとき, 垂直二等分面は, 次の方程式で表される.

$$|x_i - 0| = |x_i - 2a|$$

この方程式が意味を持つのは,

$$x_i = a \quad (0 < x_i < 2a)$$

の場合のみである.

L_∞ 距離の場合 2 点

$$\mathbf{o} = (0, 0, \dots, 0), \mathbf{q} = (q_1, q_2, \dots, q_d) \quad (q_i = 0 \text{ or } 2a)$$

を考える. このとき, $\mathbf{x} = (x_1, x_2, \dots, x_d) \quad (\forall i, 0 < x_i < a)$ という点が L_∞ 距離で, $\mathbf{o}$ に近いことを示す.

まず, L_∞ 距離 dist_∞ を計算する.

$$\begin{aligned} \text{dist}_\infty(\mathbf{o}, \mathbf{x}) &= \max_{i=1, \dots, d} \{|x_i|\} < a. \\ \text{dist}_\infty(\mathbf{q}, \mathbf{x}) &= \max_{i,j} \{|x_i|, |2a - x_j|\} \\ &= \max_j \{|2a - x_j|\} > a. \end{aligned}$$

ただし, i は座標値が 0 である座標の添え字, j は座標値が $2a$ である座標の添え字とする. また, $0 < x_i < a$ から, 最後の等号が成り立ち, $|2a - x_j| > a$ となる. また, $\text{dist}_\infty(\mathbf{o}, \mathbf{x}) (< a)$ は, $\text{dist}_\infty(\mathbf{q}, \mathbf{x}) (> a)$ よりも小さいことがわかる. よって, $\mathbf{x}$ は, $\mathbf{q}$ よりも $\mathbf{o}$ に近いことが示された. つまり, $\{\mathbf{x} \mid 0 < x_i < a, i = 1, \dots, d\}$ となる領域は $\mathbf{o}$ のボロノイ領域の一部である. もし, すべての j で $x_j = a$ となると,

$$\text{dist}_\infty(\mathbf{o}, \mathbf{x}) = \text{dist}_\infty(\mathbf{q}, \mathbf{x}) = a$$

となり, 境界となる. ただし, 他の x_i は, $0 < x_i < a$ の条件を満たすとする.

対称性から, $-a < x_i < 0$ となる座標軸がある場合も, 同様に示すことが出来る. よって,

$$\{\mathbf{x} \mid |x_i| < a, i = 1, \dots, d\}$$

が $\mathbf{o}$ のボロノイ領域に含まれ, $|x_i| = a$ が境界となる. □

上の補題と, 定理 1 から次の定理を示すことができる.

定理 2 正方格子配置に対する絶対近傍を考える. このとき, 絶対近傍の全体により全ての空間が被覆される.

証明: 補題から, 正方格子配置に対する L_1 距離によるボロノイ領域と L_∞ 距離によるボロノイ領域が一致することが示された. また, 絶対近傍がこの 2 つのボロノイ領域の交わりとなるので, L_1 距離 (L_∞ 距離) でのボロノイ領域がそのまま絶対近傍となる. 全ての絶対近傍を考えると, L_1 距離でのボロノイ図となるので, 全ての領域が被覆される. □

面心格子配置

次に、面心格子配置というものを考える。面心格子配置を以下で定義する。

$$\left\{ (n_1, n_2, \dots, n_d) \mid \sum_i n_i = (2\mathbb{Z} + 1)a, n_i = \mathbb{Z}a \right\}.$$

面心格子配置の絶対近傍を図 6 に示した。

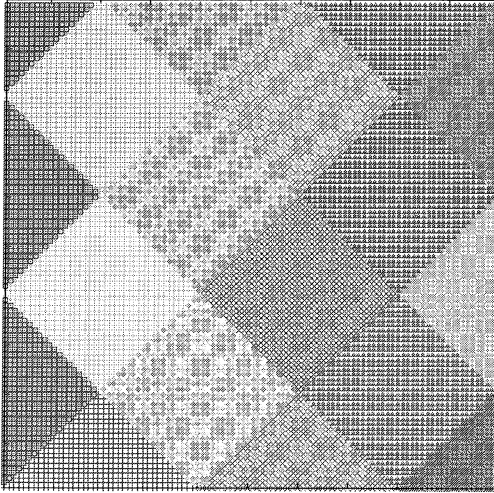


図 6 面心格子配置の絶対近傍

2 次元での被覆率は、ほぼ 1 であった。3 次元 32 点の被覆率が 0.9991, 4 次元 40 点での被覆率が 0.9981 となった。

ここでは、2 次元の場合に、面心格子配置の絶対近傍で全平面が被覆されることを示す。

定理 3 2 次元面心格子配置の絶対近傍の全体は、平面をすべて被覆する。

証明: Lee [3] では、2 次元での L_p 距離での垂直二等分線が明示されている。Lee らの図の一部を書き直したものが図 7 である。この図は、 q_i, q_j を結ぶ線分 $\overline{q_i q_j}$ の傾きが ± 1 ときの L_p 距離での垂直二等分線を表している。この図から以下のことがわかる。

L_1 距離 $\overline{q_i q_j}$ を対角線とする正方形の内部では、その正方形のもう一つの対角線が垂直二等分線となる。左上、右下の斜線の領域も q_i, q_j から等

距離な点の集合である。これらの領域を以下では**等距離領域**と呼ぶ。

L_p 距離 ($p = 2, 3, \dots, \infty$) 任意の p に対して、ユークリッド距離 (L_2) の場合と同じ垂直二等分線になる。

2 番目の事実から、 $p = 2, 3, \dots, \infty$ の場合は、垂直二等分線が一致することがわかる。

以下では、面心格子に対する L_1 距離によるボロノイ図を考えると、等距離領域は表れず、対角線部分のみが現れることを示す。

$q_i(a, 0), q_j(2a, a)$ とし、 $q_k(a, 2a), q_l(0, a)$ という 2 点を加えて考える (図 8 参照)。これら 4 点を結んだ四角形のコピーにより平面を被覆することができるので、この四角形内で等距離領域が出てこなければ、平面上のどこにも出てこず、対角線部分のみがボロノイ図の境界として残ることがわかる。

四角形 $q_i q_j q_k q_l$ 内のボロノイ領域を考える。 q_i, q_l の垂直二等分線 $B_1(q_i, q_l)$ は、原点と (a, a) を通る線分と等距離領域とからなる。 $B_1(q_i, q_j)$ の左上の等距離領域は、この二等分線の q_k 側にあるため、 q_i, q_j よりも q_k に近いことがわかる。同様に q_l にも近いことがわかる (図 8 参照)。よって、これら 4 点の垂直二等分線を描くと、等距離領域が他の点に近くなるため、ボロノイ領域の境界は、点 (a, a) を通る傾き ± 1 の直線と四角形 $q_i q_j q_k q_l$ の交わりである。この境界は、 L_p の場合と一致している。

以上から、任意の L_p において、ボロノイ領域の境界が一致したので、絶対近傍の全体が平面を被覆する。 □

5 まとめ

本稿では、ランダムな点と格子点を母点とした絶対近傍の被覆率について述べた。ランダムな点からなる母点集合によりデータ空間を分割する場合、2 次元で約 0.8, 4 次元で約 0.6 という被覆率が得られた。これらの被覆率は、母点集合の点数やデータの取りうる最大値によりあまり変化していない。つまり、絶対近傍の占める領域の全空間に対する比率は、母点集合にほとんど依存しないということがわかつ

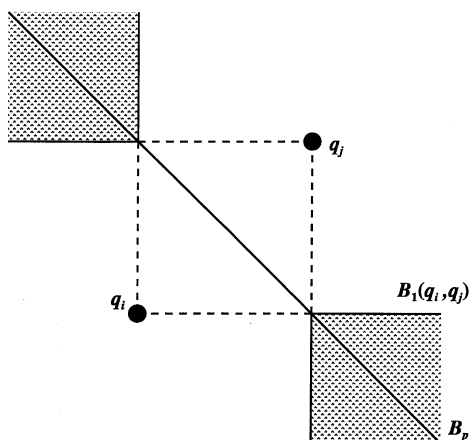


図7 2次元での L_p 距離での垂直二等分線 B_p

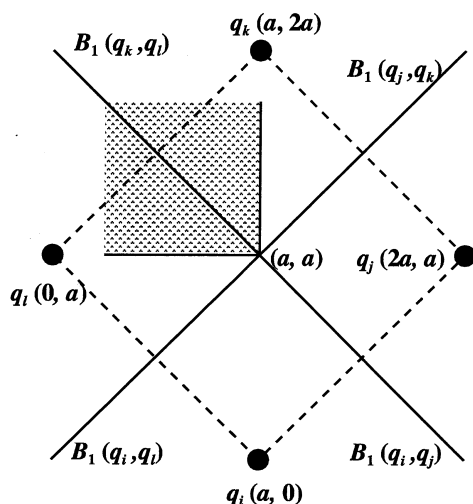


図8 面心格子をなす4点と L_1 距離での垂直二等分線

た. このことから, 母点集合の点数を気にすることなく, 母点集合をとり, 空間の一部の被覆として用いることができる. また, 空間全体を被覆していないため, 空間の分割として用いるには, 母点集合を多重化し, 空間全体を被覆するなどの手法が必要となる. この際にどの程度多重化しなければ, 空間全体を被覆できないかという指針も得られた.

また, 格子点に対する実験から, 次のことがわ

かった. 被覆率という観点からは正方格子を用いることがもっともよい. これは, バケット法が索引構造のための空間分割には適していることを理論的, 実験的に示したことになる. d 次元空間での格子点からそれぞれの領域の最遠点までの L_1 距離, L_∞ 距離を考える. 格子点から最遠点まで距離は, L_1 距離で da , L_∞ 距離で a となる. この値を双方とも持つと, 検索時の絞込みはほとんど働かないことになる. つまり, 正方格子は, mm-GNAT を用いた任意の L_p 距離での検索にはあまり向いていないことがわかる. 一方, 面心格子では, 一般の次元では被覆率は 1.0 には達しないまでもランダムな点からなる母点集合の場合よりも高い値を持つ. また, 格子点から最遠点までの距離は, L_1 距離で a , L_∞ 距離でも a と同じ値を持つため, mm-GNAT には適した点集合であることがわかる.

今後は, ランダム, 正方格子, 面心格子からなる点集合を用いて, 空間分割を行い, どの程度 mm-GNAT の検索性能が向上するのかを調べていきたい.

参考文献

- [1] C. Böhm, S. Berchtold and A. D. Keim: Searching in High-Dimensional Spaces: Index Structures for Improving the Performance of Multimedia Databases, *ACM Computing Surveys*, Vol. 33, No. 3, 2001, pp. 322–373.
- [2] E. Chávez, G. Navarro, R. Baeza-Yates and J. L. Marroquin: Searching in Metric Spaces, *ACM Computing Surveys*, Vol. 33, No. 3, 2001, pp. 273–321.
- [3] D. T. Lee: Two-Dimensional Voronoi Diagrams in the L_p -Metric, *Journal of ACM*, Vol. 27, No 4, 1980, pp. 604–618.
- [4] A. Okabe, B. Boots, K. Sugihara and S. N. Chiu: Spatial Tessellations: Concepts and Applications of Voronoi Diagrams, Second Edition, John Wiley & Sons Ltd, 2000.
- [5] K. Onishi: Intersection of Voronoi Regions

- by L_p distance, Proc. of Japan Conference on Discrete and Computational Geometry '99, Tokai University, Tokyo, 1999, pp. 26-28.
- [6] K. Onishi, M. Kobayakawa and M. Hoshi: mm-GNAT: index structure for arbitrary L_p norm, Proc. of The Second IEEE International Workshop on Multimedia Databases and Data Management (MDDM2007), Istanbul, Turkey, 2007, *to appear*.