

ランダム複体とパーシステントホモロジー

九州大学・マス・フォア・インダストリ研究所 白井朋之・平岡裕章

Tomoyuki Shirai and Yasuaki Hiraoka

Institute of Mathematics for Industry, Kyushu University

1 はじめに

図 1 は (正方形内の) ポアソン点過程のサンプルの各点に半径 t の円板をあわせて描き t を大きくしていったもののスナップショットである。つまり、点配置 $\xi = \sum_i \delta_{x_i}$ に対して、 $\xi(t) = \bigcup_i B(x_i, t)$ を対応させることにより、 $\mathbb{R}^2$ の部分集合に値をとる増加過程が定義される。まず連結性に注目すると、 $t = 0$ では



図 1: ポアソン点 20 個のフィルトレーション

各点がすべて連結成分であるが、 t が大きくなるに従い異なる連結成分が合体していき、連結成分の個数は減少していく。ホモロジー群の言葉で言えば 0-次元のホモロジー群 $H_0(\xi(t))$ の次元 (ランク) が t に関する減少関数であると言える。これは geometrical percolation の問題を考えていると言ってよいだろう。

次に、(ちょっとわかりにくい) 右から 5 番目の図で中央右下に穴が一つ発生している。右から 4 番目の図ではその穴の他に二つの穴が発生している。右から 3 番目の図では三つあった穴が一つになり、その一つは最後の図まで持続 (persistent) している。もう少し t が大きくなるとこの穴もつぶれる。ここで述べた穴の数に関することは 1-次元のホモロジー群 $H_1(\xi(t))$ を調べることによりわかる。 $H_1(\xi(t))$ を調べることはパーコレーションの問題のある種の高次元版を考えていると言ってよいだろう。

このように、 $\{\xi(t)\}_{t \geq 0}$ に対して、 $H_0(\xi(t))$, $H_1(\xi(t))$ を調べれば各 t 毎の $\xi(t)$ の位相的な様子がわかる。パーシステントホモロジーはこの考えをもう少し押し進めて、各 t 毎のスナップショットの穴の様子のみならず、(時間) パラメータ $t \geq 0$ 全体での穴 (ホモロジー群の元) の生成・消滅を理論的に扱うための道具として考案された。つまり、過程 $\{\xi(t)\}_{t \geq 0}$ に対するある種のホモロジー群 $H_i(\{\xi(t)\}_{t \geq 0})$ を考えるのである。最近では、topological data analysis (TDA) をキーワードに様々な研究が活発になされている分野である。

最後に、パーシステントホモロジーとは一見無関係に見える最小全域木の問題について述べる。 n 点の完全グラフ $K_n = (V_n, E_n)$ の各辺に独立に $[0, 1]$ 上の一様確率変数 $\{U_e, e \in E_n\}$ (ウェイト) を与える。Cayley の定理によると K_n 内にある全域木の個数は n^{n-2} 個である。各全域木 T 上のウェイトの和 $U_T := \sum_{e \in T} U_e$ を T の総ウェイトと呼び、 U_T の最小全域木全体にわたる最小値を $L_0 := \min_T U_T$ と定義する。このとき、Frieze(1985) によって、 $\mathbb{E}[L_0] \rightarrow \zeta(3)$ となることが示されている ([3])。最小全域木自体はランダムでないウェイトでももちろん定義できる。最小全域木を探索するアルゴリズムに Kruskal algorithm と呼ばれるものが存在する。 n 個の孤立点からスタートし、ウェイトの小さい辺から順にサイクルができないようにグラフに加えていき、 $(n-1)$ 個の辺を加えた所で最小全域木が得られるというものである。辺のウェイトはこのアルゴリズムにおいては生成時間と呼んでも自然であろう。このように生成時間に辺を加えていく操作は、与えたウェイトが上のような一様確率変数の場合は Erdős-Rényi グラフ過程そのものである。

この報告では、最小全域木の単体複体への拡張とパーシステントホモロジーの関係や、Erdős-Rényi グラフ過程の複体への拡張とそれに対する Frieze の定理の類似について論じる。

RIMS 研究集会「確率論シンポジウム」, Dec. 16-19, 2014

本研究は科学研究費挑戦的萌芽研究 No.26610025 および基盤研究 (B) No.26287019 の助成を受けたものです。

†その辺を加えてサイクルができるときは加えない、サイクルができないときは加える。

2 単体複体のフィルトレーションとパーシステントホモロジー

2.1 単体複体

上に述べたようないわゆる geometrical graph から, Čech 複体と呼ばれるホモトピー同値な単体複体が定義できるので, 単体複体を考察することによって一般性は失なわれない.

以下, 単体複体の定義を述べる. V を有限集合とすると, $\sigma \subset V$ が $|\sigma| = k+1$ のとき, σ は k -単体 (k -simplex) もしくは k -面 (k -face) という. このとき, $\dim(\sigma) = k$ とあらわす.

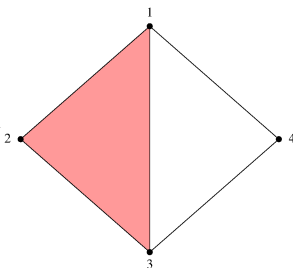
定義 2.1. X を V の非空部分集合の族とする. (V, X) が (抽象的) 単体複体であるとは条件

(1) $\{v\} \subset X$ for all $v \in V$, (2) $\sigma \in X, \tau (\neq \emptyset) \subset \sigma \implies \tau \in X$

をみたすときをいう. 条件 (2) は「 X は非空の部分集合を取る操作に関して閉じていること」を意味する.

$d = \max_{\sigma \in X} \dim(\sigma)$ のとき (V, X) を d -次元単体複体という.

以下, 簡単のため, しばしば X を単体複体という.



- $X = \{1, 2, 3, 4, 12, 13, \underline{14}, 23, \underline{34}, \underline{123}\}$ は 2 次元単体複体.
- X は facet $\{14, 34, 123\}$ (極大な単体) で定まる. 実際, 定義 2.1(2) の条件をみたすように facet の非空部分集合をすべて取ればよい.

$$X = \underbrace{\underbrace{\{1, 2, 3, 4\}}_{X_0} \cup \underbrace{\{12, 13, 14, 23, 34\}}_{X_1}}_{X^{(1)}: \text{a graph}} \cup \underbrace{\{123\}}_{X_2}$$

ここで, $X_k = \{\sigma \in X : \dim(\sigma) = k\}$ は X の k -face の全体とし, $X^{(k)} = \{\sigma \in X : \dim(\sigma) \leq k\}$ を X の k -スケルトンという.

注意. 1 次元単体複体はグラフであるから, 1-スケルトンは単体複体のベースになっているグラフである.

2.2 d -Linial-Meshulam ランダム複体過程

$V = \{1, 2, \dots, n\}$ に対して, Δ_{n-1} を V を facet とする $(n-1)$ -次元単体複体とする. つまり, $\Delta_{n-1} = 2^V \setminus \{\emptyset\}$ である. $(\Delta_{n-1})_d$ は上の定義より V の d -単体の全体, つまり V の $(d+1)$ -点部分集合の全体である. 各 $\sigma \in (\Delta_{n-1})_d$ に対して, 独立な一様確率変数 $t(\sigma) \in [0, 1]$ を考えて,

$$X^{(d)}(t) = (\Delta_{n-1})^{(d-1)} \sqcup \{\sigma \in (\Delta_{n-1})_d : t(\sigma) \leq t\}$$

とおくと, 各 d 毎に, $\{X^{(d)}(t), 0 \leq t \leq 1\}$ は単調非減少な d -次元単体複体に値をとる確率過程となる. 各 t 毎定義されるランダム複体は Linial-Meshulam ([8]) によって詳しく考察されているので, ここでは d -Linial-Meshulam ランダム複体過程 (以後, $\text{LM}^{(d)}$ -process) とよぶ. $d=1$ の場合は, $X^{(1)}(0) = (\Delta_{n-1})^{(0)} = V$ で, $(\Delta_{n-1})_1$ の各元が辺に対応するので, $X^{(1)}(t)$ はいわゆる Erdős-Rényi ランダムグラフ過程である.



図 2: Erdős-Rényi graph process ($d=1$)

$d = 2$ のときは, 完全グラフ $X^{(2)}(0) = (\Delta_{n-1})^{(1)}$ からスタートして, $\binom{n}{3}$ 個ある 2-face (三角形) をランダムに加えていき, $X^{(2)}(1) = (\Delta_{n-1})^{(2)}$ で終了する.

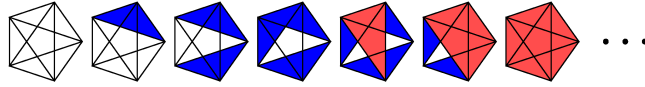


図 3: 2-Linial-Meshulam ランダム複体過程 (LM⁽²⁾-process)

2.3 単体複体のランダムフィルトレーション

単体複体 (V, X) の部分複体の増大列 $\mathbb{X} = (X(t))_{t \geq 0}$ もしくは $\mathbb{X} = (X(t))_{t=0,1,2,\dots}$ をフィルトレーションという. 前者の連続パラメータの場合は右連続性を仮定する. つまり, $s < t$ のとき $X(s) \subset X(t)$ で, $X(t+) = \bigcap_{t < u} X(u) = X(t)$ が成り立つとする. 以後 t が連続パラメータの場合は連続時間, t が離散パラメータの場合は離散時間ということにする.

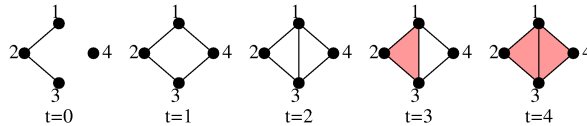


図 4: 離散時間フィルトレーションの例

定義 2.2 (生成時間 (birth time)). フィルトレーション $\mathbb{X} = (X(t))_{t \in \mathbb{T}}$ が与えられたとき,

$$T(\sigma) = t \quad \text{if} \quad \begin{cases} \sigma \in X(t) \setminus X(t-1) & (X(-1) = \emptyset), \quad \mathbb{T} = \{0, 1, 2, \dots\} \\ \sigma \in X(t+) \setminus X(t) & \mathbb{T} = [0, \infty) \end{cases}$$

を単体 σ の生成時間という.

注意 2.3. 2.2 節で定義した LM^(d)-process は $(\Delta_{n-1})^{(d)}$ のランダムな連続フィルトレーションを定義しており, $\{t(\sigma) : \sigma \in (\Delta_{n-1})_d\}$ は σ のランダムな生成時間に対応する. また, 定義より LM^(d)-process において $(d-1)$ 次元以下の単体 $\eta \in (\Delta_{n-1})_{d-1}$ の生成時間は $t(\eta) = 0$ である.

2.4 パーシステントホモロジー

単体複体 (V, X) の部分複体のフィルトレーション $\mathbb{X} = (X(t))_{t=0,1,2,\dots}$ が与えられたとする. ここでは, 簡単のために離散時間フィルトレーションを考えることにする. 通常ホモロジー群の定義と同様に, 鎖群, 境界準同型, パーシステントホモロジー群という順番で定義していく. パーシステントホモロジーに関する詳しい内容は [6](離散時間フィルトレーションの場合), 連続時間フィルトレーションの場合の定義などについては [5] を参照のこと.

定義 2.4 (k -鎖群 $C_k(\mathbb{X})$). $\mathbb{F}$ は体とする. フィルトレーション $\mathbb{X}$ をもつ単体複体に対する k -鎖群を

$$C_k(\mathbb{X}) := \bigoplus_{t=0}^{\infty} C_k(X(t)), \quad C_k(X(t)) := \left\{ \sum_{\sigma \in X(t)_k} a_{\sigma} \langle \sigma \rangle : a_{\sigma} \in \mathbb{F} \right\}$$

とする。ここで、 $\langle \sigma \rangle$ は向き付けられた単体、 $\mathbb{F}$ -係数多項式環 $\mathbb{F}[x]$ の $C_k(\mathbb{X})$ への作用は

$$x \cdot (c_0, c_1, c_2, \dots) = (0, c_0, c_1, c_2, \dots)$$

を自然に線形に拡張して定義する[†]。 $x \cdot C_k(X(t)) \subset C_k(X(t+1))$ となるので、 $C_k(\mathbb{X})$ は次数付き $\mathbb{F}[x]$ -加群となる。

写像 $\iota_t : C_k(X(t)) \rightarrow C_k(\mathbb{X})$ を $\sigma \mapsto (0, 0, \dots, 0, \overbrace{\sigma}^t, 0, 0, \dots)$ と定義する。この写像を用いて

$$\Xi_k := \{e_\sigma := \iota_{T(\sigma)}\sigma, \sigma \in X_k := \bigoplus_{t \geq 0} (X(t))_k\}$$

と定義すると、 $C_k(\mathbb{X})$ の次数付き $\mathbb{F}[x]$ -加群としての基底を与える[‡]。例えば、図 4 の例では $C_1(\mathbb{X})$ の基底は

$$\begin{aligned} e_{12} &= (\langle 12 \rangle, 0, 0, 0, 0, \dots), & e_{13} &= (0, 0, \langle 13 \rangle, 0, 0, 0, \dots), & e_{14} &= (0, \langle 14 \rangle, 0, 0, 0, \dots), \\ e_{23} &= (\langle 23 \rangle, 0, 0, 0, 0, \dots), & e_{34} &= (0, \langle 34 \rangle, 0, 0, 0, \dots) \end{aligned}$$

で与えられ、 $\dim C_1(\mathbb{X}) = 5$ である。

定義 2.5 (境界作用素). $\partial_k(x) : C_k(\mathbb{X}) \rightarrow C_{k-1}(\mathbb{X})$ は以下のように定義される。

$$\partial_k(x)e_\sigma := \sum_{j=0}^k (-1)^j x^{T(\sigma) - T(\sigma_j)} e_{\sigma_j}.$$

σ_j は $\sigma = \langle v_0 \dots v_k \rangle$ のとき $\sigma_j = \langle v_0 \dots \widehat{v_j} \dots v_k \rangle$ をあらわす。

$\mathbb{R}$ 上の多項式環 $\mathbb{R}[x]$ で考えると、図 4 の例では

$$\begin{aligned} \partial_2(x)e_{123} &= \partial_2(x)((0, 0, 0, \langle 123 \rangle, 0, \dots)) = (0, 0, 0, \partial_2 \langle 123 \rangle, 0, \dots) = (0, 0, 0, \langle 12 \rangle - \langle 13 \rangle + \langle 23 \rangle, 0, \dots) \\ &= x^3(\langle 12 \rangle, 0, 0, 0, 0, \dots) - x(0, 0, \langle 13 \rangle, 0, 0, 0, \dots) + x^3(\langle 23 \rangle, 0, 0, 0, 0, \dots) \\ &= x^3 e_{12} - x e_{13} + x^3 e_{23}. \end{aligned}$$

同様にして他の基底に関しても計算すると、境界作用素の表現行列は以下のように与えられる。基底をあらわす行列の添字は e_σ を $\langle \sigma \rangle$ とあらわしている。

$$\partial_1(x) = \begin{array}{c} C_0 \setminus C_1 \\ \begin{matrix} \langle 1 \rangle \\ \langle 2 \rangle \\ \langle 3 \rangle \\ \langle 4 \rangle \end{matrix} \end{array} \begin{array}{ccccc} \langle 12 \rangle & \langle 13 \rangle & \langle 14 \rangle & \langle 23 \rangle & \langle 34 \rangle \\ \begin{bmatrix} -1 & -x^2 & -x & & \\ 1 & & & -1 & \\ & x^2 & & 1 & -x \\ & & x & & x \end{bmatrix} \end{array}, \quad \partial_2(x) = \begin{array}{c} C_1 \setminus C_2 \\ \begin{matrix} \langle 12 \rangle \\ \langle 13 \rangle \\ \langle 14 \rangle \\ \langle 23 \rangle \\ \langle 34 \rangle \end{matrix} \end{array} \begin{array}{cc} \langle 123 \rangle & \langle 134 \rangle \\ \begin{bmatrix} x^3 & \\ -x & x^2 \\ & -x^3 \\ x^3 & \\ & x^3 \end{bmatrix} \end{array}$$

$x=1$ とおいた $\partial_1(1), \partial_2(1)$ はいわゆる (有向) 接続行列 (incidence matrix) に一致する。

ホモロジー群の場合と同様に $\partial_{k-1}(x) \circ \partial_k(x) = 0$ がなりたつことが簡単にチェックできるので、

$$B_k(\mathbb{X}) := \text{Im } \partial_{k+1}(x) \subset Z_k(\mathbb{X}) := \ker \partial_k(x)$$

がわかる。

定義 2.6. k 次のパーシステントホモロジー群を $H_k(\mathbb{X}) := Z_k(\mathbb{X})/B_k(\mathbb{X})$ と定義する。

[†] x の作用は時間のシフトに対応。

[‡]体を係数とする多項式環が単項イデアル整域 (PID) であることが本質である。整数係数の多項式環 $\mathbb{Z}[x]$ は PID ではない。

一般に有限生成次数付き $\mathbb{F}[x]$ 加群の構造定理は以下のように与えられる.

定理 2.7. $\mathbb{F}$ を体とし, $M = \bigoplus_{i \geq 0} M_i$ は次数付き $\mathbb{F}[x]$ -加群とする. このとき,

$$M \cong \bigoplus_{i=1}^p ((x^{b_i})/(x^{d_i})) \oplus \bigoplus_{i=p+1}^{p+q} (x^{b_i}).$$

ただし, $b_i, d_i \in \mathbb{Z}_{\geq 0}$, $b_i < d_i$ である.

よって, 定理 2.7 をパーシステントホモロジー群に適用すれば,

$$H_k(\mathbb{X}) \cong \bigoplus_{i=1}^p ((x^{b_i})/(x^{d_i})) \oplus \bigoplus_{i=p+1}^{p+q} (x^{b_i})$$

を得る. b_i, d_i はそれぞれ i 番目の k 次パーシステントホモロジー類の生成時間と消滅時間をあらわす. このとき,

$$\underbrace{\text{フィルトレーション } \mathbb{X}}_{\text{input}} \Rightarrow \underbrace{\{(b_i, d_i), i = 1, 2, \dots, p\} \cup \{(b_i, \infty), i = p+1, p+2, \dots, p+q\}}_{\text{output}}$$

という図式を得られることが重要である. (b_i, d_i) に対して, $\ell_i := d_i - b_i$ を生存時間という. 上の場合, $p+1 \leq i \leq p+q$ の場合は $\ell_i = \infty$ と理解する. 例えば, $\text{LM}^{(d)}$ -process では確率 1 で $q=0$ となるので, $\ell_i, i=1, 2, \dots, p$ はすべて有限である. (実際, 定義より $d_i \leq 1$ となる.) このとき, $L_k = \sum_{i=1}^p \ell_i$ を k 次パーシステントホモロジー類の生存時間の和という. この $\{(b_i, d_i)\}_i$ の情報をパーシステント図やバーコードなど視覚的にあらわすと便利である. これらの図としての極限定理も興味深い問題である. 極限定理を考えるには, 例えばパーシステント図ならば点過程とみなすのが自然である ([5]).

3 最小全域木とその高次元への拡張

グラフ $G = (V, E)$ 上の部分グラフ (V, S) が G の spanning tree (全域木) であることは, 次のようにして特徴付けができる. (i) サイクルをもたない, (ii) 連結, (iii) $|S| = |V| - 1$. これらの性質を単体的複体に拡張することを考える.

定義 3.1. X を単体複体として, $k \leq \dim X$ とする. $S \subset X_k$ に対して k -次元部分複体 $X_S = S \sqcup X^{(k-1)}$ を定義する. $S \subset X_k$ が k -spanning acycle もしくは k -次元 spanning acycle であるとは条件

$$(a) \tilde{H}_k(X_S) = 0, \quad (b) |\tilde{H}_{k-1}(X_S)| < \infty \quad (\tilde{H}_* \text{ は被約ホモロジー})$$

をみたすときをいう. X の k -次元 spanning acycles の全体を $S^{(k)}$ とあらわす.

注意 3.2. (1) この定義は $X = \Delta_{n-1}$ の場合に G.Kalai によって与えられたものである [7]. グラフ ($k=1$) のときは, (i) は $\tilde{H}_1(X_S) = 0$ に対応し条件 (a) と同じであるが, (ii) は $\tilde{H}_0(X_S) = 0$ となるが対応する条件 (b) は $k \geq 2$ の場合は $\tilde{H}_{k-1}(X_S) = 0$ ではなく「ねじれ部分」をゆるすものとする. この定義の下, Kalai は完全グラフ上の全域木の数に関する Cayley の定理を spanning acycle に対して拡張している. (2) 定義の中に全域木の場合の (iii) に相当する条件がないが, (a) と (b) の条件から $|S|$ の条件が従う.

定理 3.3 ([5]). X は有限な単体複体で $\tilde{\beta}_{d-1}(X^{(d)}) = \tilde{\beta}_{d-2}(X^{(d-1)}) = 0$ ($1 \leq d \leq \dim X$) をみたすとする. このとき, $\mathbb{X} = \{X(t)\}_{t \in [0, \infty]}$ を X の連続フィルトレーションとし, L_k を k 次パーシステントホモロジーの生存時間の和とすると以下の等式が成り立つ.

$$L_{d-1} = \min_{T \in S^{(d)}} \text{wt}(T) - \max_{S \in S^{(d-1)}} \text{wt}(X_{d-1} \setminus S) \quad (3.1)$$

$$= \int_0^\infty \tilde{\beta}_{d-1}(t) dt. \quad (3.2)$$

ここで, $S \subset X^{(k)}$ に対して $\text{wt}(S) = \sum_{\sigma \in S} T(\sigma)$, $\tilde{\beta}_i(t)$ は複体 $X(t)$ の i 次元被約 Betti 数である.

注意 3.4. (1) (3.1) は 1 節で述べた Kruskal algorithm の単体複体版に相当する内容である。

(2) (3.2) はパーシステントホモロジーの情報 $\{(b_i, d_i)\}_i$ を i 番目の客の来店時間・退出時間とみなすと待ち行列における Little の公式と解釈することができる。実際、生存時間 $\ell_i = d_i - b_i$ を客の店内滞在時間、時刻 t における Betti 数を店内の客の数 $\tilde{\beta}_{d-1}(t) = \sum_{i=1}^{N(t)} I(b_i \leq t < d_i)$, $N(t)$ は t 以前に来た客の数 (時刻 t 以前に生成する生成元の数) とすると以下のように解釈できる。

$$\underbrace{\frac{1}{T} \int_0^T \tilde{\beta}_{d-1}(s) ds}_{\text{平均店内人数}} = \frac{1}{T} \sum_{i=1}^{N(T)} (d_i - b_i) = \underbrace{\frac{N(T)}{T}}_{\text{客の到着率}} \cdot \underbrace{\frac{1}{N(T)} \sum_{i=1}^{N(T)} \ell_i}_{\text{客の平均滞在時間}} = \frac{1}{T} L_{d-1}$$

4 主結果とその証明のスケッチ

定理 4.1. $d \in \mathbb{N}$ とする。 L_{d-1} を n 点上の $\text{LM}^{(d)}$ -process の $(d-1)$ 次パーシステントホモロジーの生存時間の和とする。このとき、 $0 < c < C$ が存在して、 $n \rightarrow \infty$ のとき

$$cn^{d-1} \leq \mathbb{E}[L_{d-1}] \leq Cn^{d-1} \log \log n \quad (4.1)$$

となる。

注意 4.2. シンポジウムの講演の際には上の定理のように、上界と下界が一致していなかったが、下界が正しい漸近形であることが最近示された [5]。つまり $\mathbb{E}[L_{d-1}] = O(n^{d-1})$ である。また、極限值 $\lim_{n \rightarrow \infty} n^{-(d-1)} \mathbb{E}[L_{d-1}]$ についても議論している。これは、 $d=1$ のときは 1 節で述べた Frieze の最小全域木に関する定理の拡張となっている。詳細については [5] を参照せよ。

4.1 定理 4.1 の証明のスケッチ

下からの評価には (3.1)、上からの評価には (3.2) を用いる。

(i) 下からの評価。注意 2.3 で述べたように、 $\text{LM}^{(d)}$ -process $\{X^{(d)}(t)\}$ では任意の $(d-1)$ -face η に対し $t(\eta) = 0$ であるから、(3.1) より S を minimum spanning $(d-1)$ -acycle とすると $L_{d-1} = \text{wt}(S)$ となる。さらに $L_{d-1} = \text{wt}(S) \geq \sum_{i=1}^{|S|} u_i$ である。ここで、 $0 \leq u_1 \leq u_2 \leq \dots \leq u_{\binom{n}{d}} \leq 1$ は $\{t(\sigma), \sigma \in (\Delta_{n-1})_d\}$ を小さい順に並べたものである。順序統計量の平均の公式と $|S| = \binom{n-1}{d}$ であることより

$$\mathbb{E}[L_{d-1}] \geq \sum_{i=1}^{|S|} \mathbb{E}[u_i] = \sum_{i=1}^{|S|} \frac{i}{\binom{n}{d+1} + 1} = O(n^{d-1}).$$

(ii) 上からの評価。Betti 数の単体数による自明な評価 $\tilde{\beta}_{d-1}(t) \leq f_{d-1}(t) \leq \binom{n}{d}$ と (3.1) を使う。 $\text{LM}^{(d)}$ -process では $\tilde{\beta}_{d-1}(t)$ は t に関して単調減少であることに注意して、 $\tau_{d-1} = \inf\{t \geq 0 : \tilde{\beta}_{d-1}(t) = 0\}$ とおくと、

$$\mathbb{E}[L_{d-1}] = \mathbb{E} \left[\int_0^{\tau_{d-1}} \tilde{\beta}_{d-1}(t) dt \right] \leq \binom{n}{d} \mathbb{E}[\tau_{d-1}] = O(n^{d-1} \log n).$$

しかし、この評価では $n^{d-1} \log \log n$ までは出ない。

最後の等式に用いる $\mathbb{E}[\tau_{d-1}] \sim O(\frac{\log n}{n})$ の評価には大道具を使っており、4.3 節で簡単に触れる。

(iii) 上からの評価の改善。4.2 節でみるように、離散モース理論と同値な概念である acyclic matching とよばれる部分マッチングを用いたホモトピー変形を行なうことにより、考察対象の複体の単体数を減らすことができる。この後、上と同様な自明な評価を行うことにより、(ii) で用いた Betti 数の自明な評価を改善することができ、 $\log n$ を $\log \log n$ に落せる。

4.2 Acyclic matching と離散モース理論

定義 4.3. (1) (V, X) を単体複体とする. 分割 $X = A \sqcup Q \sqcup K$ と全単射 $\varphi: Q \rightarrow K$ が存在して, 各 $\sigma \in Q$ に対して $\sigma \subset \varphi(\sigma)$ が余次元 1, つまり $|\sigma| = |\varphi(\sigma)| - 1$ となるとき, $\mathcal{M} = (A, \varphi: Q \rightarrow K)$ を部分マッチングという. また, A に属する単体を critical な単体とよぶ.

(2) 単体の列 $\sigma_0 \subset \tau_0 \subset \sigma_1 \subset \tau_1 \subset \sigma_2 \subset \cdots \subset \sigma_m \subset \tau_m \subset \sigma_0$ が各 $i = 0, \dots, m$ に対して $\tau_i = \varphi(\sigma_i)$ となるとき, closed $\mathcal{M}$ -path という. $\varphi(\sigma_0) = \tau_0 = \tau_1 = \varphi(\sigma_1)$ であるから $m \geq 2$ であることを注意しておく.

(3) 部分マッチング $\mathcal{M}$ が *acyclic matching* であるとは, closed $\mathcal{M}$ -path が存在しないときをいう.

このとき以下の定理が知られている.

定理 4.4 ([2]). X は acyclic matching $\mathcal{M}$ をもつ単体複体とすると, X は各 critical k -単体に対してそれぞれ k -cell を対応させることによって得られる CW-複体にホモトピー同値となる.

つまりこの定理により, よい ($|A|$ が小さくなるような) acyclic matching を構成できれば, 先の Betti 数の自明な評価は, $\tilde{\beta}_k \leq |A|$ に改善することができる.

$\text{LM}^{(d)}$ -process に対して acyclic matching $(A, \varphi: Q \rightarrow K)$ を以下で構成しよう. $\text{LM}^{(d)}$ -process ではすべての $(d-1)$ -単体は含まれているので $X^{(d)}(t)$ の $(d-1)$ -スケルトンは可縮である. acyclic matching を $(d-1)$ -単体と d -単体のレベルで構成する. これを $(d-1, d)$ -type acyclic matching とよぶことにする.

1. V に適当に全順序 $1 <_{\text{lex}} 2 <_{\text{lex}} \cdots <_{\text{lex}} |V|$ を入れる. 辞書的順序により自然に X にも全順序が入る. この順序で σ が τ より前にあらわれるときは, $\sigma <_{\text{lex}} \tau$ とあらわす.
2. $\sigma \in X_{d-1}$ に対して, $\tau = \text{lexmin}\{\tilde{\tau} \in X_d \mid \sigma \subset \tilde{\tau}, \sigma <_{\text{lex}} \tilde{\tau}\}$ となるような $\tau \in X_d$ が存在するとき, ペアリングとして $\varphi: \sigma \mapsto \tau$ をつけくわえる.
3. これを繰り返して最終的に残る単体をすべて A とする.

これが acyclic matching になっていることは簡単にチェックできる. また, この構成は一般の単体複体にも適用可能である. 定理 4.4 を $\text{LM}^{(d)}$ -process とここで構成した $(d-1, d)$ -type acyclic matching に適用する.

定理 4.1 の証明のスケッチ. $\text{LM}^{(d)}$ -process の時刻 t における単体複体 $X^{(d)}(t)$ に上で定義した $(d-1, d)$ -type acyclic matching を構成して, $f_{d-1}^*(t)$ を critical $(d-1)$ -cells の数とする. 定理 4.4 より $\tilde{\beta}_{d-1}(t) \leq f_{d-1}^*(t)$ を得る. $i_1 i_2 \dots i_{d-1} j$ が critical であることと, 任意の $k \geq j+1$ に対して $i_1 i_2 \dots i_{d-1} k$ があらわれないことは同値であることに注意して, 平均 $\mathbb{E}[f_{d-1}^*(t)]$ を計算すると

$$\mathbb{E}[f_{d-1}^*(t)] = \sum_{\sigma \in (\Delta_{n-1})_{d-1}} \mathbb{P}(\sigma \text{ is critical}) = \sum_{j=d}^n \sum_{1 \leq i_1 < i_2 < \cdots < i_{d-1} < j} \mathbb{P}(i_1 i_2 \dots i_{d-1} j \text{ is critical}).$$

よって, $\mathbb{P}(\sigma \in X^{(d)}(t)) = t \ (\forall \sigma \in \Delta_{n-1}^{(d)})$ に注意すれば

$$\mathbb{E}[\tilde{\beta}_{d-1}(t)] \leq \mathbb{E}[f_{d-1}^*(t)] = \sum_{j=d}^n \binom{j-1}{d-1} (1-t)^{n-j}.$$

十分大きい $C > 0$ に対して $a_n = \frac{C \log n}{n}$ において, (3.2) を用いて上の不等式を変形していくと,

$$\begin{aligned} \mathbb{E}[L_{d-1}] &= \mathbb{E} \left[\int_0^{\tau_{d-1}} \tilde{\beta}_{d-1}(t) dt \right] \leq \mathbb{E} \left[\int_0^{a_n} \tilde{\beta}_{d-1}(t) dt \right] + \binom{n}{d} \mathbb{P}(\tau_{d-1} > a_n) \\ &\leq \int_0^{a_n} \sum_{j=d}^n \binom{j-1}{d-1} (1-t)^{n-j} dt \leq \binom{n-1}{d-1} \sum_{k=1}^{n-d+1} \frac{1}{k} \{1 - (1-a_n)^k\} \\ &\leq \binom{n-1}{d-1} \left\{ \sum_{k=1}^{\lfloor a_n^{-1} \rfloor} a_n + \sum_{k=\lfloor a_n^{-1} \rfloor + 1}^n \frac{1}{k} \right\} \leq \binom{n-1}{d-1} \{1 + (\log n - \log \lfloor a_n^{-1} \rfloor)\} = O(n^{d-1} \log \log n). \end{aligned}$$

を得る. $\lesssim$ には次節の命題 4.8 を用いる. □

4.3 コホモロジー消滅とスペクトルギャップ

最後に $\text{LM}^{(d)}$ -process の Betti 数に関する hitting time $\tau_{d-1} = \inf\{t \geq 0 : \tilde{\beta}_{d-1}(t) = 0\}$ の評価についてコメントする. この評価については二つの道具を用いる. 単体複体 X に対して, 単体 $\tau \in X$ の X におけるリンク (link) を $lk_X(\tau) = \{\sigma \setminus \tau : \tau \subseteq \sigma\}$ と定義する. リンク $lk_X(\tau)$ はまた単体複体になり, 特に X が d -次元で τ が $(d-2)$ -単体のとき, リンク $lk_X(\tau)$ はグラフとなる. しばしば単体複体の問題をグラフの問題に帰着させる有力な手段となる. ここでは以下の注意が重要な役割をはたす.

注意 4.5. n 点上の $\text{LM}^{(d)}$ -process の $(d-2)$ -face τ のリンク $\{lk_{X^{(d)}(t)}(\tau)\}_{t \in [0,1]}$ は $(n-d-1)$ 点上の Erdős-Rényi graph process $\{G(n-d-1, t)\}_{t \in [0,1]}$ と同分布になる.

定理 4.6 (Garland, Ballman-Swiatkowski [1]). X が pure d -次元の単体複体とする[†]. 各 $(d-2)$ -face τ のリンク $lk_X(\tau)$ (グラフ) が連結で, そのラプラシアン²のスペクトルギャップが $\lambda_2(lk_X(\tau)) > 1 - \frac{1}{d}$ となるとき, $H^{d-1}(X, \mathbb{Q}) = 0$ である. 特に, $\tilde{\beta}_{d-1}(X) = 0$ となる.

注意 4.5 より, 定理 4.6 の条件は $G(n-d-1, t)$ の連結性とスペクトルギャップの問題に帰着される. $G(n-d-1, t)$ の連結性とスペクトルギャップの問題に帰着される. 後者については例えば Hoffman-Kahle-Paquette [4] による以下の結果を利用する.

定理 4.7 (Hoffman-Kahle-Paquette [4]). $t = t(n) = (\frac{1}{2} + \delta) \frac{\log n}{n}$ ($\delta > 0$) に対し, $G(n, t)$ を Erdős-Rényi graph process とする. $\mu = (n-1)t$ (平均次数) とおく. $\tilde{G} = (\tilde{V}, \tilde{E})$ を $G \sim G(n, t)$ の巨大連結成分とするとき, 任意の $\epsilon > 0$ に対して正定数 $A(\delta, \epsilon) > 0$ が存在して

$$\mathbb{P}(\lambda_2(\tilde{G}) \leq 1 - \frac{A}{\sqrt{\mu}}) \leq A \left(ne^{-(2-\epsilon)\mu} + e^{-\mu^{1/4} \log n} \right). \quad (4.2)$$

が成り立つ. ここで, $\tilde{n} = |\tilde{V}|$, $\lambda_2(\tilde{G})$ は $\tilde{G}$ 上の SRW に対するラプラシアン²のスペクトルギャップである. これより, 任意の $\epsilon > 0$ と $M > 0$ に対してある正定数 $C > 0$ が存在して, $G(n, \frac{C \log n}{n})$ の巨大連結成分上 $\mathbb{P}(\lambda_2(\tilde{G}) \leq 1 - \epsilon) = o(n^{-M})$ となる.

定理 4.6 と定理 4.7 などから以下の結果を得る. (本質的に [4] など得られている.)

命題 4.8. $\text{LM}^{(d)}$ -process に対して $\tau_{d-1} = \inf\{t \geq 0 : \tilde{\beta}_{d-1}(t) = 0\}$ とする. このとき, 任意の $\alpha > 0$ に対して $C > 0$ が存在して,

$$\mathbb{P}(\tau_{d-1} > \frac{C \log n}{n}) = o(n^{-\alpha}). \quad (4.3)$$

また, $n \rightarrow \infty$ で $\mathbb{E}[\tau_{d-1}] = O(\frac{\log n}{n})$ が成り立つ.

参考文献

- [1] W. Ballman, J. Swiatkowski. On L^2 -cohomology and property (T) for automorphism groups of polyhedral cell complexes. *Geom. Funt. Anal.* 7 (1997), 615–645.
- [2] R. Forman. Morse theory for cell complexes. *Adv. Math.* 134 (1998), 90–145.
- [3] A. M. Frieze. On the value of a random minimum spanning tree problem. *Discrete Applied Math.* 10 (1985), 47–56.
- [4] C. Hoffman, M. Kahle and E. Paquette. Spectral Gaps of Random Graphs and Applications to Random Topology. *Discrete Math.* 309 (2009), 1658–1671.
- [5] Y. Hiraoka and T. Shirai, Minimum spanning acycle and lifetime of persistent homology in the Linial-Meashulam process, <http://arxiv.org/abs/1503.05669>
- [6] 平岡裕章, タンパク質構造とトポロジー——パーシステントホモロジー群入門——, 共立出版, 2013 年.
- [7] G. Kalai. Enumeration of Q -acyclic simplicial complexes. *Israel J. Math.* 45 (1983), 337–351.
- [8] N. Linial and R. Meshulam. Homological connectivity of random 2-complexes. *Combinatorica* 26 (2006), 475–487.

[†]単体複体が pure d -次元であるとは, すべての facet が d -次元であるときをいう.