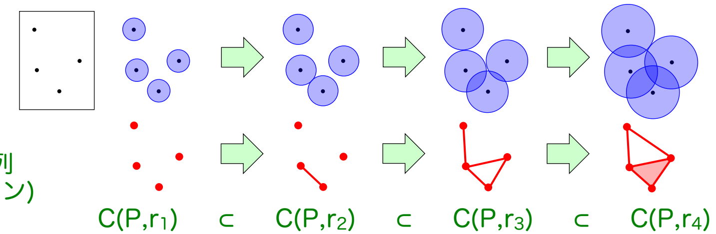
イントロダクション：「パーシステント図」とは？

これは位相的データ解析（TDA）における道具であり、以下のようにして構築される.

step①：各点を太らせていき、単体的複体の増大列（フィルトレーション）を作る.



step②：適当な体$\mathbb{K}$上で鎖複体を作り i 次ホモロジー群をとる.

クイバーA_nの表現　$H_i(C(P,r_1)) \to H_i(C(P,r_2)) \to H_i(C(P,r_3)) \to H_i(C(P,r_4))$

クイバーとは有向グラフのことで $A_n := \underset{1}{\bullet} \to \underset{2}{\bullet} \dashrightarrow \underset{n-1}{\bullet} \to \underset{n}{\bullet}$ である. 一般にクイバーQに対し各頂点に有限次元ベクトル空間、各矢に線形写像を割り当てたものを「Qの表現」という.

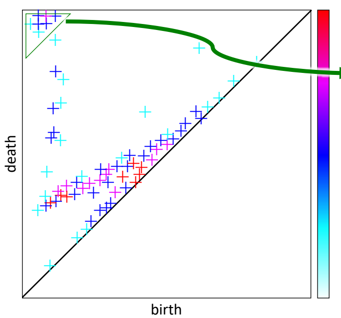
こうして得られた表現を「パーシステンス表現」といい
とくにこの例のようにA_nの表現になっているものを「1Dパーシステンス表現」という.

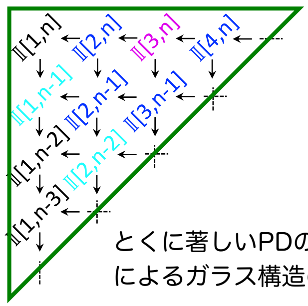
実は任意の表現は直既約表現の直和のかたちに（順序と同型を除き）一意的に分解できる.
直既約表現の同型類の完全代表系を $\mathcal{L}$ とおく. $Q = A_n$ の場合は $\mathcal{L}$ を次のように取れる！

$$\mathcal{L} = \{\, \mathbb{I}[b,d] := \underset{\text{1-st}}{0} \dashrightarrow 0 \xrightarrow{0} \underset{\text{b-th}}{\mathbb{K}} \xrightarrow{1} \mathbb{K} \dashrightarrow \underset{\text{d-th}}{\mathbb{K}} \xrightarrow{0} 0 \dashrightarrow \underset{\text{n-th}}{0} \mid b, d \in \{1,2,\dots,n\},\ b \le d \,\}$$

表現 M に、直既約因子として $L \in \mathcal{L}$ がいくつ含まれているかを $d_M(L)$ と書くことにする.
「パーシステント図」とは、この写像 $d_M : \mathcal{L} \to \mathbb{Z}_{\ge 0}$ をプロットした図のことである.

step③：パーシステンス表現の直既約分解を求め、その結果をビジュアライズする.





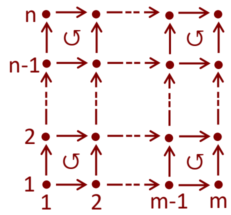
PDから何が分かる？

因子 $\mathbb{I}[b,d]$ の存在は
時刻 b で生じた「穴」が
時刻 d で埋まったという
事象に対応している.

とくに著しいPDの応用例として、平岡・中村らによるガラス構造の特徴付け(2016)が知られる！

モチベーション：PDをパワーアップしたい…！しかし…！

ポイントクラウドの構造が時系列でどう変化するのかも調べたい！
そのためにグリッド $G_{m,n}$ の表現（2Dパーシステンス表現）に対し同様の計算をしたい！ つまり与えられた2Dパーシステンス表現Mに対する写像d_Mを計算したい！ しかしこれは事実上不可能に近い.



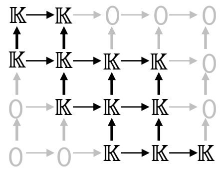
グリッド$G_{m,n}$

一般に $G_{m,n}$ の表現の全体は「野性的」と呼ばれ、手に負えない！
始域 $\mathcal{L}$ は無限集合になり、極めて多種多様な要素から構成される.

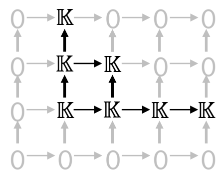
ではどうするか？：「区間表現」という良い直既約表現に注目！

区間表現とは1Dパーシステンス表現の直既約因子がもっていた性質を拡張した概念で
$G_{m,n}$ の区間表現は具体的には次のような右下がりの階段型（staircase）の表現になる！
この全体を $\mathbb{I}$ とおく. これの特徴付けなどに関する詳細は arXiv:1812.05261 を参照.

$G_{5,4}$の区間表現と、その記述の例　（$\mathbb{K}$と$\mathbb{K}$の間は恒等写像、その他は零写像）



$\{\,[1,2]_4,\ [1,4]_3,\ [2,4]_2,\ [3,5]_1\,\}$



$\{\,[2,2]_4,\ [2,3]_3,\ [2,5]_2\,\}$

この $\mathbb{I}$ は有限集合になってくれる. そこで、与えられた2Dパーシステンス表現Mを
$\mathbb{I}$ の元の足し合わせとして近似的に分解することを考える！ これを「区間近似」と呼ぶ.
多元環の表現論と組合せ論の知見を駆使していくことで、次のような定義が導かれる！

Definition (区間近似)

$M \in \operatorname{rep} G_{m,n}$に対し、写像 $\partial^*_M : \mathbb{I} \to \mathbb{Z}$ を次で定める.

$$\partial^*_M(I) := \sum_{S \subseteq \operatorname{cov} I} (-1)^{\#S} d_{\operatorname{comp}^*_{\vee S} M}(\operatorname{comp}^*_{\vee S} \vee S)$$

ここで $\operatorname{comp}^*_{\vee S} M$ は「圧縮」という操作＊によって得られるMよりも“小さな”表現である.
この圧縮についての詳細は、向かいのスライド（講演）か arXiv:1911.01637 を参照！

区間近似の有用性：区間近似からどんな情報が取り出せるのか？

区間近似は、元の表現の「繋がりの情報」である rank を完全に保存している！

Theorem（区間近似は、元の表現の射のrankの情報を保つ！）

$M \in \operatorname{rep} G_{m,n}$とする. $G_{m,n}$上の任意の道 $s \dashrightarrow t$ に対して次が成り立つ！

$$\operatorname{rank} M_{s \dashrightarrow t} = \sum_{J \in \mathbb{I}} \partial^*_M(J) \cdot \operatorname{rank} J_{s \dashrightarrow t}$$

さらに、区間近似には次のような特長がある！

Theorem（区間近似がもつ性質）

[1]: 任意の $M \in \operatorname{add} \mathbb{I}$ に対して $\partial^*_M = d_M$
[2]: 任意の $M, N \in \operatorname{rep} G_{m,n}$ に対して $\partial^*_{M \oplus N} = \partial^*_M + \partial^*_N$

この性質から、Mが実際に直既約因子にもつ区間表現はちゃんと感知できているといえる.
現実のポイントクラウドデータ由来のパーシステンス表現には、非区間表現はほとんど含まれない可能性も示唆されており、データ解析における区間近似の有用性が期待される.

Let’s click! ⇒ このアルゴリズムのデモプログラムで遊ぶ！