

СИБИРСКИЕ ЭЛЕКТРОННЫЕ МАТЕМАТИЧЕСКИЕ ИЗВЕСТИЯ

Siberian Electronic Mathematical Reports

<http://semr.math.nsc.ru>

Том 11, стр. 130–141 (2014)

УДК 519.6+515.146

MSC 55-04

УСТОЙЧИВОСТЬ ИНТЕГРАЛЬНЫХ ПЕРСИСТЕНТНЫХ ДИАГРАММ

А.Е. АБЖАНОВ, Я.В. БАЗАЙКИН

АБСТРАКТ. We define concept of integral persistent diagram which involve geometrical characteristics of excursion sets and prove stability of such diagrams.

Keywords: computational topology, persistence, stability.

1. ВВЕДЕНИЕ

Персистентные диаграммы, введенные в [1] (см. также [2]) являются важнейшим инструментом для вычислительной топологии, позволяющим, например, получить качественную информацию о «топологической динамике» множеств экскурсии $\{f \leq a\}$ функции f , заданной на некотором топологическом пространстве. С точки зрения приложений, функцию f можно интерпретировать как некоторое поле той или иной природы, и качественное описание множеств экскурсии дает представление о характере и свойствах этого поля. В такой ситуации естественно предполагать, что функция является результатом некоторых измерений и задана лишь с точностью до некоторой погрешности. Это ставит вопрос о устойчивости топологических характеристик экскурсии к возмущениям исходного поля.

В [3] доказана теорема стабильности: если заданы две функции f и g на топологическом пространстве X , то расстояние между их персистентными диаграммами можно оценить через L_∞ -норму их разности:

$$d_B(D(f), D(g)) \leq \|f - g\|_\infty.$$

ABZHANOV, A.E., BAZAIKIN, YA.V., STABILITY OF INTEGRAL PERSISTENCE DIAGRAMS.

© 2014 Абжанов А.Е., Базайкин Я.В.

Первый автор поддержан Министерством образования и науки Республики Казахстан (программа 055 «Научная и/или научно-техническая деятельность»); второй автор поддержан РФФИ (гранты 12-01-00873-а и 12-01-92104-ЯФ-а).

Поступила 18 сентября 2013 г., опубликована 17 февраля 2014 г.

Эта теорема является обоснованием использования персистентности в прикладных задачах.

В данной работе мы изучаем вопрос о модификации определения персистентной диаграммы с тем, чтобы, она учитывала не только чисто топологическую информацию о динамике множеств экскурсии, но и геометрические соотношения, которым эта динамика подчиняется. Мы предлагаем следующую конструкцию. Рассмотрим динамику изменения компонент связности множеств экскурсии. В стандартном определении персистентных 0-гомологий каждая точка (p, q) в персистентной диаграмме добавляется в момент p рождения новой компоненты связности, а координата q отвечает времени ее слияния с более старшей компонентой, что формулируется как правило «старший выживает». Мы заменяем это правило на правило «большой выживает», сравнивая некоторую геометрическую характеристику сливающихся компонент, называемую интегральным объемом. Полученную диаграмму мы обозначаем $ID(f)$.

Теорема. Пусть X — полиэдр с мерой μ и f, g — непрерывные кусочно-линейные функции на нем. Тогда

$$d_B(ID(f), ID(g)) \leq 8\mu(X)\|f - g\|_\infty.$$

Значение теоремы заключается в том, что определенная нами интегральная персистентная диаграмма является устойчивой по отношению к возмущениям исходной функции, что обосновывает ее потенциальное использование.

2. ПЕРСИСТЕНТНЫЕ ДИАГРАММЫ

В этом параграфе мы напомним основные сведения об устойчивости персистентных диаграмм, следуя [3]. Пусть X — топологическое пространство, f — непрерывная функция на нем. Для $a \in \mathbb{R}$ введем обозначение: $X_a = f^{-1}(-\infty, a]$. Число $a \in \mathbb{R}$ называется гомологическим критическим значением, если для некоторого k гомоморфизм, индуцированный вложением $f_* : H_k(X_{a-\varepsilon}) \rightarrow H_k(X_{a+\varepsilon})$ не является изоморфизмом для любого достаточно малого $\varepsilon > 0$ (группы гомологий рассматриваются с коэффициентами в \mathbb{Z}_2). Непрерывную функцию f называют ручной, если у нее существует лишь конечное число гомологических критических значений, группы гомологий $H_k(X_a)$ конечно порождены для всех k и всех $a \in \mathbb{R}$, и если отрезок $[x, y]$ не содержит гомологических критических значений ручной функции, то гомоморфизм $f_x^y : H_k(X_x) \rightarrow H_k(X_y)$, индуцированный вложением, является изоморфизмом. Например, для простоты можно считать, что X является полиэдральным пространством, а f — кусочно-линейная функция.

Предположим, что $f : X \rightarrow \mathbb{R}$ — ручная функция. Пусть $a_1 < a_2 < \dots < a_n$ — ее гомологические критические значения. Рассмотрим промежуточные значения b_0, \dots, b_n : $b_{i-1} < a_i < b_i$. Пополним выбранные точки следующим образом: $b_{-1} = a_0 = -\infty$ и $b_{n+1} = a_{n+1} = +\infty$. Для каждой пары индексов $0 \leq i < j < n + 1$ определим кратность точки (a_i, a_j) положив

$$\mu_i^j = \beta_{b_{i-1}}^{b_j} - \beta_{b_i}^{b_j} + \beta_{b_i}^{b_{j-1}} - \beta_{b_{i-1}}^{b_{j-1}},$$

где $\beta_x^y = \dim(\text{Im}(f_x^y))$.

Персистентной диаграммой $D(f) \subset \mathbb{R}^2$ функции f называется множество точек (a_i, a_j) , $i, j = 0, \dots, n + 1$ с учетом кратностей μ_i^j , в объединении с множеством точек диагонали $\Delta = \{(x, x) | x \in \mathbb{R}\}$, взятых с бесконечной кратностью.

Замечательным свойством персистентной диаграммы $D(f)$ является ее стабильность по отношению к возмущениям функции f . Для формулировки теоремы стабильности, нам потребуется еще несколько определений. Если даны две непрерывные функции f, g на X , то $\|f - g\|_\infty = \sup_{p \in X} |f(p) - g(p)|$. Далее, если $x = (x_1, x_2), y = (y_1, y_2) \in \mathbb{R}^2$, то $\|x - y\|_\infty = \max\{|x_1 - y_1|, |x_2 - y_2|\}$. Определим два типа расстояния между множествами $X, Y \subset \mathbb{R}$ (точки множеств предполагаются имеющими кратность):

$$d_H(X, Y) = \max_{x \in X, y \in Y} \left\{ \sup_x \inf_y \|x - y\|_\infty, \sup_y \inf_x \|y - x\|_\infty \right\},$$

$$d_B(X, Y) = \inf_{\gamma} \sup_{x \in X} \|x - \gamma(x)\|_\infty,$$

где γ пробегает по всем биекциям из X в Y . Первое расстояние называют расстоянием Хаусдорфа, второе — расстоянием бутылочного горлышка. Очевидно, что $d_H \leq d_B$.

Теорема.[3]

Пусть X — триангулируемое топологическое пространство, $f, g : X \rightarrow \mathbb{R}$ — ручные функции. Тогда

$$d_H(D(f), D(g)) \leq d_B(D(f), D(g)) \leq \|f - g\|_\infty.$$

3. ИНТЕГРАЛЬНЫЕ ПЕРСИСТЕНТНЫЕ ДИАГРАММЫ

Пусть K — конечный симплициальный комплекс. Тогда его носитель $X = |K|$ будем называть полиэдром, если он наделен кусочно-линейной римановой метрикой, согласованной на общих гранях. Тем самым на X задана форма объема и мера μ . В дальнейшем будем рассматривать только непрерывные кусочно-линейные функции на X , и для каждой такой функции f положим $X_a = \{p \in X \mid f(p) \leq a\}$.

Для $Y \subset X$ через $C(Y)$ будем обозначать множество компонент связности. Для сокращения записи будем писать $C(X_a) = C_a$ для $a \in \mathbb{R}$. Для любых значений $x < y$ через $f_x^y : C_x \rightarrow C_y$ будем обозначать естественное отображение конечных множеств, индуцированное вложением $X_x \subset X_y$.

Для каждого $a \in \mathbb{R}$ и каждой компоненты связности $Y \in C_a$ определим ее интегральный объем (Рис. 1) по формуле

$$I(Y) = \mu(Y)a - \int_Y f(x)d\mu(x) = \int_Y (a - f(x))d\mu(x).$$

Пусть a_1, \dots, a_n — все критические значения функции f , дополним их значениями $a_0 = -\infty, a_{n+1} = +\infty$. Для каждого целого $i, 1 \leq i \leq n$ рассмотрим множество компонент связности $C_i = C_{a_i}$. Естественно считать, что $a_0 = \emptyset, a_{n+1} = X$.

Определим граф Γ_0 следующим образом. В качестве множества вершин рассмотрим $\cup_{i=1}^n C_i$. Каждую вершину $c \in C_i$ соединим направленным ребром $\gamma(c) = \overrightarrow{cd}$ с вершиной $d = f_i(c) = f_{a_i}^{a_{i+1}}(c)$. Ясно, что граф Γ_0 является деревом. На вершинах графа Γ_0 введем отношение порядка: будем считать $c \leq d$, если существует путь в графе Γ_0 , состоящий из направленных ребер, начинающийся в c и заканчивающийся в d .

Нам потребуется дополнительная структура на графе Γ_0 . Для того, чтобы ее задать, рассмотрим его произвольную вершину $d \in C_i$. Пусть $\emptyset \neq f_{i-1}^{-1}(d) =$

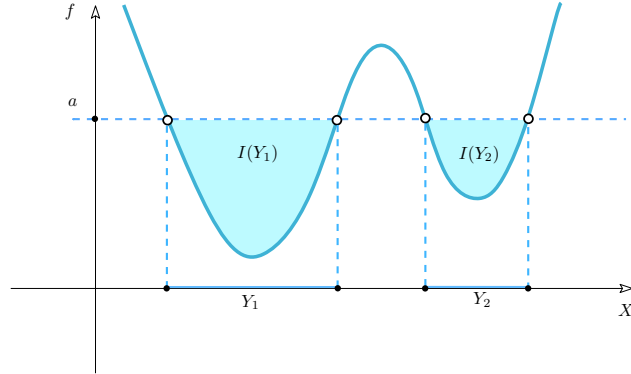


Рис. 1. Интегральный объем компоненты связности.

$\{c_1, \dots, c_k\} \subset C_{i-1}$. Положим

$$I_j^-(d) = \lim_{a \rightarrow a_i - 0} I(f_{a_{i-1}}^a(c_j)),$$

где $j = 1, \dots, k$. Ясно, что

$$\sum_{j=1}^k I_j^-(d) = I(d).$$

Будем называть интегральным объемом (либо просто — объемом) ребра $\overrightarrow{c_j d}$, $j = 1, \dots, k$ величину

$$V(c_j, d) = I_j^-(d) - I(c_j).$$

Далее, назовем весом ребра $\overrightarrow{c_j d}$ величину

$$w(c_j, d) = I(d) - I_j^-(d),$$

где $j = 1, \dots, k$. Определим интегральный объем (или просто — объем) каждой вершины графа Γ_0 следующим образом. Если (в прежних обозначениях) $f_{i-1}^{-1}(d) = \emptyset$, т.е. у вершины d отсутствуют входящие ребра, то положим $V(d) = 0$. В противном случае, положим

$$V_j^-(d) = V(c_j) + V(c_j, d), V(d) = \max_{j=1, \dots, k} V_j^-(d)$$

Для удобства в дальнейшем будем считать множество $f_{i-1}^{-1}(d) = \{c_1, \dots, c_k\}$ упорядоченным таким образом, что $V_1^-(d) \geq V_2^-(d) \geq \dots \geq V_k^-(d)$, т.е. $V(d) = V_1^-(d)$. Весом вершины c назовем сумму

$$w(c) = \sum_{d \geq c} w(\gamma(d)).$$

Очевидно, что веса вершин графа могут быть найдены рекуррентно, начиная от $c \in C_n$. Вершинам множества C_{n+1} припишем вес 0.

Теперь определим интегральную персистентную диаграмму $ID(f)$. Для этого (в прежних обозначениях) образуем новый граф Γ следующим образом. Для каждой вершины d , такой что $f_{i-1}^{-1}(d) \neq \emptyset$ отъединим ребра $\overrightarrow{c_j d}$, $j = 2, \dots, k$ от графа Γ_0 таким образом, что каждое такое ребро будет оканчиваться новой

вершиной d_j , $j = 2, \dots, k$, а ребро $\overrightarrow{c_1 d}$ будет оканчиваться «старой» вершиной d .

Понятно, что граф Γ состоит из несвязного объединения связных подграфов, каждый из которых представляет собой линейную последовательность вершин и звеньев между ними. Каждому такому подграфу сопоставим пару вершин: начальную $c \in C_k$ и конечную, которая совпадает либо с вершиной из C_{n+1} , либо с одной из вершин вида d_j , $d \in C_i$. В первом случае мы добавляем к персистентной диаграмме точку $(w(c), \infty)$, во втором — точку $(w(d_j) + I(d) - V_j^-(d), w(d_j) + I(d))$. Множество всех таких точек в \mathbb{R}^2 , с учетом их кратностей, объединенное с множеством точек диагонали Δ с бесконечной кратностью мы будем называть интегральной персистентной диаграммой, и обозначать $ID(f)$.

Замечание 1. Можно охарактеризовать принцип построения интегральной персистентной диаграммы как «выживает больший»: если двигаться по дереву Γ_0 снизу вверх, то выживает та ветка, объем которой больше. При этом в последующем объем отрезанной ветки не учитывается и не суммируется в ветках, стоящих в дереве выше. Вместо диаграммы можно рассмотреть баркод, который изоморфен Γ . При этом длины отрезков в баркоде в точности отвечают объемам умирающих компонент. Расположение отрезков на координатной оси при этом меняется в соответствии с заданными выше весами. Рис. 2 иллюстрирует разницу между классическим и интегральным баркодами.

Введем следующее обозначение, которое будет нам полезно в дальнейшем. Для каждой вершины Y графа $\Gamma_0(f)$, $\Gamma(f)$ (соответственно, $\Gamma_0(g)$, $\Gamma(g)$) положим $f(Y) = a$ ($g(Y) = b$), если $Y \in C(f^{-1}(-\infty, a])$ ($Y \in C(g^{-1}(-\infty, b])$).

Чтобы прояснить смысл интегральной персистентной диаграммы, построим вспомогательную функцию \tilde{f} на X следующим образом. Пусть $p \in X$, $f(p) = a$. Предположим, что $a_{i-1} \leq a < a_i$, $i < 1$. Выберем компоненту связности Y множества X_a , в которой лежит точка p и рассмотрим $c = (f_{a_{i-1}}^a)^{-1}(Y)$. Положим

$$\tilde{f}(p) = I(Y) + w(c).$$

Лемма 1.

Функция \tilde{f} непрерывна и $ID(f) = D(\tilde{f})$.

Доказательство. Пусть сначала $a_{i-1} < f(p) = a < a_i$. Рассмотрим компоненту $d \in C_i$, содержащую p . Пусть $f_{i-1}^{-1}(d) = \{c_1, \dots, c_k\}$, где $c_j \in C_{i-1}$, $j = 1, \dots, k$. Без ограничения общности, можно считать, что $p \in f_{a_{i-1}}^a(c_1)$, и при этом $p \notin f_{a_{i-1}}^a(c_j)$ для $j = 2, \dots, k$. Рассмотрим «срезающую» функцию $i: \mathbb{R} \rightarrow \mathbb{R}$, заданную соотношением

$$i(x) = \frac{1}{2}(|x| + x).$$

Очевидно, что $i(x) - i(y) \leq i(x - y)$ для всех $x, y \in \mathbb{R}$. Тогда

$$\tilde{f}(p) = \int_d i(f(p) - f(x)) d\mu(x) + w(c).$$

Последняя формула влечет непрерывность \tilde{f} , поскольку в рассмотренной ситуации вес $w(c)$ является константой в окрестности точки p .

Пусть теперь $f(p) = a_i$. Рассмотрим компоненту связности Y множества X_{a_i} , содержащую точку p . Пусть $f_{i-1}(Y) = \{c_1, \dots, c_k\}$. Предположим, что

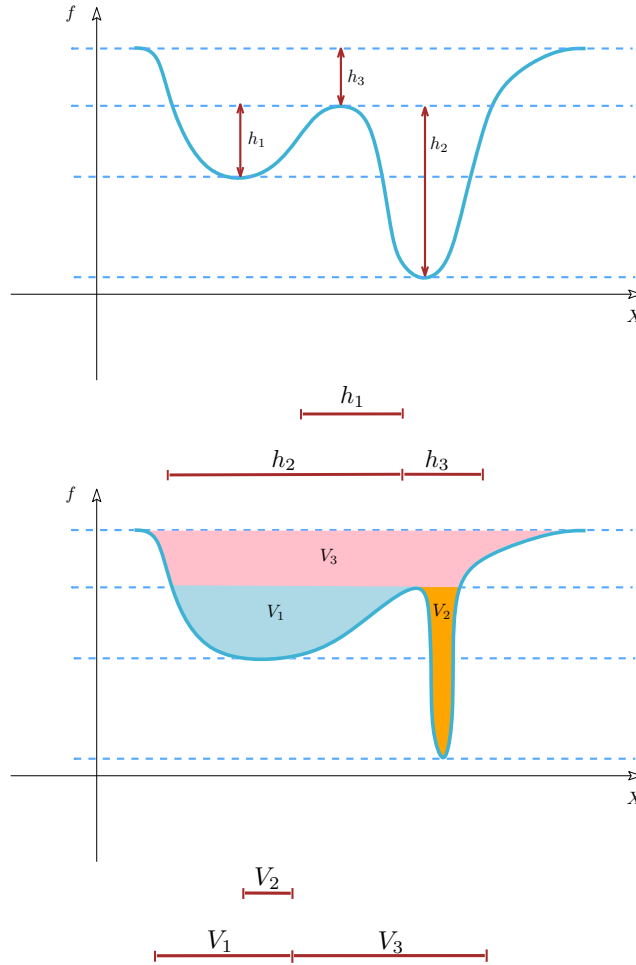


Рис. 2. Баркод, отвечающий классической (сверху) и интегральной (снизу) персистентным диаграммам.

последовательность p_m сходится к точке p при $m \rightarrow \infty$. Разделим последовательность p_m на две подпоследовательности p_m^- и p_m^+ , такие, что $f(p_m^-) \leq a_i$ и $f(p_m^+) \geq a_i$. В свою очередь, подпоследовательность p_m^- разделим на подпоследовательности $p_{j,m}^-$, $j = 1, \dots, k$, такие, что $p_{j,m}^- \in f_{a_{i-1}}^{b_{j,m}}(c_j)$ для всех m , где $b_{j,m} = f(p_{j,m}^-)$. Каждая из полученных подпоследовательностей либо конечна, либо сходится к p . Для доказательства непрерывности функции \tilde{f} в точке p нам достаточно показать, образы относительно \tilde{f} каждой бесконечной из рассмотренных подпоследовательностей сходятся к $\tilde{f}(p)$.

Действительно, в силу непрерывности f можно считать, что $f(p_m^+) < a_{i+1}$ для всех m . Обозначим $Z = f_i(Y)$. Тогда

$$\tilde{f}(p_m^+) = \int_Z i(f(p_m^+) - f(x))d\mu(x) + w(Y) \rightarrow_{m \rightarrow \infty}$$

$$\int_Z i(f(p) - f(x))d\mu(x) + w(Y) = \int_Y (f(p) - f(x))d\mu(x) + w(Y) = \tilde{f}(p).$$

Далее, для всех $j = 1, \dots, k$ имеем:

$$\begin{aligned} \tilde{f}(p_{j,m}^-) &= \int_{f_{a_{i-1}}^{b_{j,m}}(c_j)} i(f(p_{j,m}^-) - f(x))d\mu(x) + \sum_{l \neq j} I_l^-(Y) + w(Y) \xrightarrow{m \rightarrow \infty} \\ I_j^-(Y) + \sum_{l \neq j} I_l^-(Y) + w(Y) &\xrightarrow{m \rightarrow \infty} \int_Y (f(p) - f(x))d\mu(x) + w(Y) = \tilde{f}(p). \end{aligned}$$

Второе утверждение леммы практически очевидно: при построении графа Γ из графа Γ_0 принцип выживания ребра с большим интегральным объемом превращается в принцип «выживания старшего» для функции \tilde{f} . Лемма доказана.

Следующая лемма вместе с леммой 1 и теоремой о стабильности классических персистентных диаграмм доказывает нашу основную теорему.

Лемма 2.

Пусть f, g — непрерывные кусочно-линейные функции на X . Тогда

$$\|\tilde{f} - \tilde{g}\|_\infty \leq 8\mu(X)\|f - g\|_\infty.$$

Доказательство. Пусть, как и ранее, $\varepsilon = \|f - g\|_\infty$. Пусть $\Gamma_0(f), \Gamma(f)$ (соответственно, $\Gamma_0(g), \Gamma(g)$) — графы, построенные по функциям f (соответственно, g) как описано в параграфе 3. Добавим к графу $\Gamma(f)$ континуальное множество вершин, причем каждая вершина v из этого множества однозначно определяется своим значением $f(v)$, которое пробегает от минимума до максимума значений функции f на X . Полученный граф обозначим $\bar{\Gamma}(f)$. Аналогично определяется $\bar{\Gamma}(g)$.

Рассмотрим число $\delta > 0$ настолько малое, что расстояние между двумя различными критическими значениями функции f и функции g больше, чем δ .

Наша ближайшая цель — построение биекции $\phi : \bar{\Gamma}(f) \rightarrow \bar{\Gamma}(g)$, которая будет отождествлять «близкие» компоненты связности. Сначала предположим, что p и q — точки минимума на X функций f и g , соответственно. Тогда p и q являются вершинами графов $\Gamma(f)$ и $\Gamma(g)$ и мы положим $\phi(p) = q$. При этом, $f(p) - g(q) \leq f(q) - g(q) \leq \varepsilon$ и $g(q) - f(p) \leq g(p) - f(p) \leq \varepsilon$, то есть $|f(p) - g(q)| \leq \varepsilon$. Положим $\Gamma'(f)$ — связный подграф в $\Gamma(f)$, содержащий все вершины d , для которых $d \geq p$. Аналогично определяется $\Gamma'(g)$.

Далее, среди всех линейных подграфов в $\Gamma(f)$, пересекающихся с $\Gamma'(f)$ только по своей вершине, выберем подграф максимальной высоты (предположим также, что его высота $> 2\varepsilon$). Пусть он начинается в точке $p_1 \in X$ и заканчивается в вершине $d \in \Gamma'(f)$. Пусть $f(d) = a$. Пусть Y — компонента связности множества $\{f \leq a - \delta\}$, содержащая точку p_1 . Теперь в графе $\Gamma(g)$ рассмотрим максимальный по высоте линейный подграф, начинающийся в некоторой точке $q_1 \in X$, заканчивающийся в точке $c \in \Gamma'(g)$ и такой, что $p_1 \in c$. Пусть $g(c) = b$ и рассмотрим компоненту связности Z множества $\{g \leq b - \delta\}$, содержащую точку q_1 . Положим $\phi(p_1) = q_1$ и $\phi(d) = c$. Так как $p_1, q_1 \in Z$, то $g(q_1) - f(p_1) \leq g(p_1) - f(p_1) \leq \varepsilon$. Покажем, что $q_1 \in Y$. Действительно, существует непрерывный путь $x(t)$ в X такой, что $g(x(t)) \leq g(p_1)$, $x(0) = p_1$,

$x(1) = q_1$ (т.к. q_1 строго ниже p_1 по линейному подграфу). В этом случае $f(x(t)) \leq g(x(t)) + \varepsilon \leq g(p_1) + \varepsilon \leq f(p_1) + 2\varepsilon < a$, т.к. мы предполагаем, что высота линейного подграфа больше 2ε . Итак, p_1 и q_1 лежат в одной компоненте связности Y . Но тогда $f(p_1) \leq f(q_1)$ и мы имеем $f(p_1) - g(q_1) \leq f(q_1) - g(q_1) \leq \varepsilon$, т.е. $|f(p_1) - g(q_1)| \leq \varepsilon$.

Покажем, что уровни a и b также отличаются не более чем на ε . Действительно, пусть $b < a - \varepsilon$. Существует непрерывный путь $x(t)$ в X такой, что $g(x(t)) \leq b$ и $x(0) = q$, $x(1) = q_1$. Тогда $f(x(t)) \leq g(x(t)) + \varepsilon \leq b + \varepsilon < a$ — противоречие, т.к. $q_1 \in Y$ и $q \notin Y$. Точно так же (симметрично по f, g) доказывается невозможность $a < b - \varepsilon$. Итак, $|a - b| \leq \varepsilon$. Теперь изменим $\Gamma'(f)$, положив его равным прежнему подграфу $\Gamma'(f)$, объединенному с найденным подграфом, начинающимся в p_1 и заканчивающимся в c ; аналогично для g . Таким образом, мы получаем индуктивный процесс, который обрывается тогда, когда в обоих графах вне $\Gamma'(f)$ (и вне $\Gamma'(g)$) остаются только линейные подграфы высоты меньшей 2ε . Их мы отображаем в добавленные подграфы нулевой высоты.

Итак, мы построили биекцию ϕ , причем для любого максимального связного подграфа в $\Gamma(f)$, его образ при ϕ имеет начало и конец, отличающийся не более чем на ε . Пусть \vec{cd} — некоторое ребро графа $\Gamma(f)$. Определим редукцию f_* функции f относительно ребра f следующим образом. Пусть $f(d) = a$. Для любого $t \in [a - \delta, a)$ рассмотрим компоненту связности A_t множества $\{f \leq t\}$, содержащую множество c . Если $p \in A_t$ для некоторого t , то положим $f_*(p) = a$, в противном случае положим $f_*(p) = f(p)$. Несложно показать, что определенная таким образом функция f_* непрерывна и ее граф $\Gamma(f_*)$ получается из $\Gamma(f)$ отсечением ребра \vec{cd} и всех ребер и вершин из которых можно последовательно подняться к ребру \vec{cd} .

Рассмотрим теперь минимальный по высоте линейный подграф \vec{cd} графа $\Gamma(f)$. Пусть $f(d) = a$ и A — компонента связности множества $\{f \leq a - \delta\}$, содержащая множество c . Рассмотрим соответствующий линейный подграф $\vec{\phi(c)\phi(d)}$ графа $\Gamma(g)$, $g(\phi(d)) = b$ и соответствующую компоненту связности B множества $\{f \leq a - \delta\}$, содержащую $\phi(c)$. Ясно, что $c = p$ — точка минимума функции f в A , и $\phi(c) = q$ — точка минимума функции g в B . По нашему построению ϕ , мы имеем $p, q \in A \cap B$. Сравним интегральные объемы компонент A и B :

$$\begin{aligned} I_f(A) - I_g(B) &= \int_A (a - \delta - f(x)) d\mu(x) - \int_B (b - \delta - g(x)) d\mu(x) = \\ &= \int_{A \cap B} ((a - b) + (g(x) - f(x))) d\mu(x) + \int_{A \setminus B} (a - \delta - f(x)) d\mu(x) - \\ &\quad - \int_{B \setminus A} (b - \delta - g(x)) d\mu(x) \leq \\ &= 2\mu(A \cap B)\varepsilon + \int_{A \setminus B} (a - \delta - f(x)) d\mu(x) - \int_{B \setminus A} (b - \delta - g(x)) d\mu(x). \end{aligned}$$

Оценим второе и третье слагаемые (в вычислениях можно опускать δ , так как в дальнейшем мы перейдем к пределу при $\delta \rightarrow 0$). Пусть $x \in A \setminus B$. Тогда $g(x) - a \leq f(x) + \varepsilon - a \leq \varepsilon$. Пусть теперь x — точка минимума функции g на

множестве $Cl(A \setminus B)$. Предположим сначала, что $g(x) \geq a - \varepsilon$, тогда $|a - g(x)| \leq \varepsilon$. Следовательно, $|a - f(x)| \leq |a - g(x)| + |g(x) - f(x)| \leq 2\varepsilon$.

Пусть теперь $g(x) < a - \varepsilon \leq b$. Пусть B' — компонента связности множества $\{g \leq b\}$, содержащая точку x . Докажем, что x — точка минимума функции g в B' . Действительно, если существует $y \in B'$, $g(y) < g(x)$, то $f(y) \leq g(y) + \varepsilon < g(x) + \varepsilon < a$, т.е. $y \in A$, что противоречит определению x . Так как $x, p \in A$, $f(y) \leq f(x)$ и подграф \vec{cd} минимальный, то существует непрерывный путь $x(t)$ в X , такой, что $x(0) = x, x(1) = p$ и $f(x(t)) \leq f(x)$ (т.е. точки x и p не разделены никаким уровнем f в A). С другой стороны, $p \in B$ и $x \in B'$, поэтому на кривой существует некоторая точка $x(t_0) = r$, такая, что $g(r) > b$. Значит, высота компоненты, содержащей B' и сливающейся с B не превосходит величины $g(r) - g(x) \leq f(r) + \varepsilon - f(x) + \varepsilon \leq 2\varepsilon$. В частности, $g(x) \geq b - 2\varepsilon \geq a - 3\varepsilon$, откуда $|a - g(x)| \leq 3\varepsilon$ и $|a - f(x)| \leq 4\varepsilon$.

Таким образом, получаем (переходя к пределу при $\delta \rightarrow 0$):

$$\int_{A \setminus B} (a - \delta - f(x)) d\mu(x) \leq 4\mu(A \setminus B)\varepsilon$$

В силу симметрии по f и g получаем

$$I_f(A) - I_g(B) \leq 2\mu(A \cap B)\varepsilon + 4\mu(A \setminus B)\varepsilon + 4\mu(B \setminus A)\varepsilon \leq 4(\mu(A) + \mu(B))\varepsilon.$$

Далее, в случае, если $g(x) \geq a - \varepsilon$, заменим функции f и g редуцированными функциями относительно ребер, которые однозначно определяются компонентами A и B . Если $g(x) < a - \varepsilon$, то мы дополнительно редуцируем функцию g относительно ребра, отвечающего компоненте B' . В силу симметрий всех рассуждений относительно замены f и g , в первом случае мы получаем неравенство $|f_* - g_*| \leq \varepsilon$. Если же $g(y) < a - \varepsilon$, то $y \in B'$ и $f_*(y) = a$, $g_*(y) = g(r) \leq f(r) + \varepsilon \leq a + \varepsilon$ и $g(r) \geq b \geq a - \varepsilon$, т.е. снова $|f_*(y) - g_*(y)| \leq \varepsilon$.

Итак, мы перешли к новым функциям f и g , предварительно получив оценку на интегральные объемы удаленных компонент связности. Продолжая этот процесс по индукции, мы можем доказать оценку на интегральные объемы всех компонент, соответствующих друг другу при биекции ϕ .

Пусть теперь $p \in X$. Тогда значения \tilde{f} и \tilde{g} в точке p различаются на сумму интегральных объемов, отвечающих критическим значениям f и g , стоящим выше a и b . Поскольку носители этих объемов (т.е. сами компоненты) не пересекаются, то предыдущее неравенство дает нам:

$$|\tilde{f}(p) - \tilde{g}(p)| \leq 8\mu(X)\varepsilon.$$

Лемма доказана.

Замечание 2. Во многих естественных ситуациях интегральная персистентная диаграмма оказывается устойчивой по отношению к малому шуму в норме L_1 , если возмущение является неположительным (Рис. 3 и 4). При добавлении положительных пиков устойчивость не наблюдается (Рис. 5), что естественно связано с выбором знака неравенства при определении компоненты связности множества экскурсии. Мы планируем исследовать устойчивость интегральных персистентных диаграмм в L_1 норме, также как проанализировать результаты их вычислений на модельных случайных полях в отдельной статье.

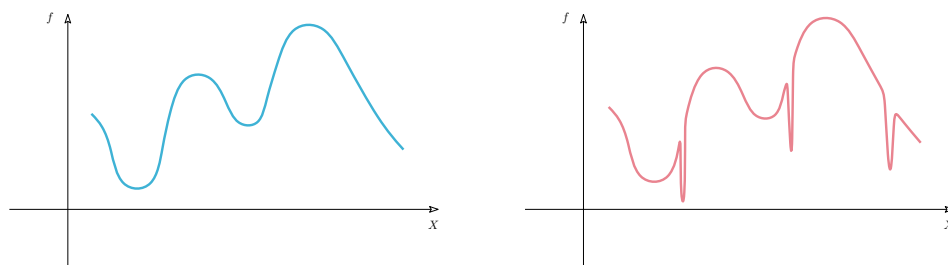


Рис. 3. Малый шум в интегральной норме.

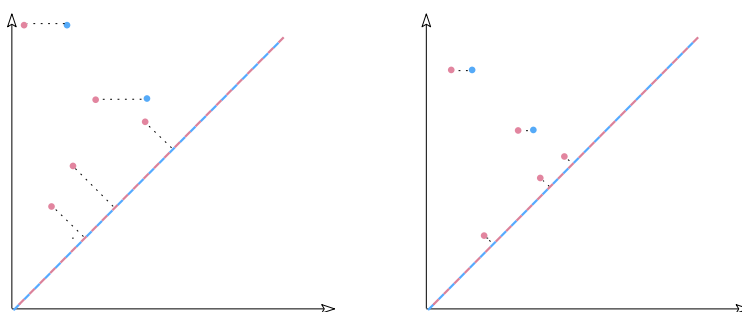


Рис. 4. Возмущение классической (слева) и интегральной (справа) персистентных диаграмм, отвечающее шуму, малому в интегральной норме.

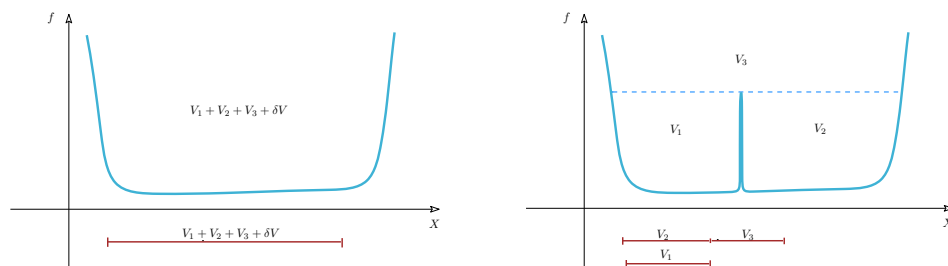


Рис. 5. Чувствительность интегральной персистентной диаграммы по отношению к «положительным» пикам: к бакоду добавляется отрезок «большой» длины.

4. АЛГОРИТМ

В данном разделе мы представляем алгоритм, вычисляющий интегральную персистентную диаграмму в случае, когда X представляет собой квадратную область. При этом мы считаем, что функция $f : X \rightarrow \mathbb{R}$ задана своими значениями в ячейках стандартной прямоугольной сетки на X , состоящей из $n \times n$ элементарных квадратов. Перед началом работы алгоритма квадраты сетки считаем упорядоченными по возрастанию значений функции f в них. Таким

образом, f_i представляет собой значение f в ячейке с номером i , $i = 1, \dots, n^2$, $f_i \leq f_{i+1}$.

Суть алгоритма состоит в следующем. Последовательно перебирая ячейки сетки в соответствии с их порядком, мы формируем множество компонент связности *componentsSet*: новая компонента добавляется, если у ячейки с номером i нет соседей, добавленных на предыдущих шагах; если соседние ячейки принадлежат ровно одной ранее добавленной компоненте, то данная ячейка поглощается этой компонентой; наконец, если существует более одной ранее добавленной компоненты, соседней с данной ячейкой, то данная ячейка поглощается компонентой с максимальным объемом, которая также поглощает остальные компоненты. В последнем случае поглощаемые компоненты погибают, а их объемы не добавляются к объему поглощающей компоненты. Объемы компонент считаются по индукции: в начале каждого цикла к объему посчитанному ранее добавляется произведение скачка $f_i - f_{i-1}$ значений функции на суммарную площадь ячеек компоненты, сформированной в конце предыдущего цикла.

Алгоритм без труда обобщается на случай трехмерных областей, или областей большей размерности.

Algorithm 1 Integral persistence diagram

```

1: componentsSet  $\leftarrow \emptyset$ 
2: for  $i := 1$  to  $n \times n$  do
3:   for each component in componentsSet do
4:     component.volume  $\leftarrow$  component.volume +  $\text{area}(\text{component.cells}) \times (f_i - f_{i-1})$ 
5:   end for
6:   neighbourComponents  $\leftarrow \emptyset$ 
7:   for each  $c_j$  in  $\text{neighbours}(c_i)$  do
8:     neighbourComponents  $\leftarrow$  neighbourComponents +  $\text{get\_component}(c_j)$ 
9:   end for
10:  if  $\text{size}(\text{neighbourComponents}) == 0$  then
11:    newComponent.cells  $\leftarrow c_i$ 
12:    newComponent.birthtime  $\leftarrow f_i$ 
13:    componentsSet  $\leftarrow$  componentsSet + newComponent
14:  else if  $\text{size}(\text{neighbourComponents}) == 1$  then
15:    component.cells  $\leftarrow$  component.cells +  $c_i$ 
16:  else if  $\text{size}(\text{neighbourComponents}) > 1$  then
17:    max\_component  $\leftarrow$   $\text{GetMaxVolumeComponent}(\text{neighbourComponents})$ 
18:    for each component in neighbourComponents that  $\neq$  max\_component do
19:      max\_component.cells  $\leftarrow$  max\_component.cells + component.cells
20:      component.deathtime  $\leftarrow f_i$ 
21:    end for
22:    max\_component.cells  $\leftarrow$  max\_component.cells +  $c_i$ 
23:  end if
24: end for
25: return componentsSet

```

СПИСОК ЛИТЕРАТУРЫ

- [1] H. Edelsbunner, D. Letscher, A. Zomorodian, *Topological persistence and simplification* // Discrete Comput. Geom., **28** (2002), 511–533. MR 1949898
- [2] G. Carlsson, A. Zomorodian, *Computing persistent homology* // In “Proc. 20th Ann. Sympos. Comput. Geom., 2004”, 347–356.
- [3] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, *Stability of Persistence Diagrams* // Discrete & Computational Geometry, **37**:1 (2007), 103–120. MR 2279866

АЙДОС ЕРАСЫЛОВИЧ АБЖАНОВ
NAZARBAYEV UNIVERSITY RESEARCH AND INNOVATION SYSTEM,
ПР. КАВАНБАЙ БАТЫРА 53,
010000, АСТАНА, РЕСПУБЛИКА КАЗАХСТАН
E-mail address: aabzhanov@nu.edu.kz

ЯРОСЛАВ ВЛАДИМИРОВИЧ БАЗАЙКИН
ИНСТИТУТ МАТЕМАТИКИ ИМ. С. Л. СОВОЛЕВА СО РАН,
ПР. АКАДЕМИКА КОПТЮГА 4,
630090, НОВОСИБИРСК, РОССИЯ
E-mail address: bazaikin@math.nsc.ru