Правительство Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего профессионального образования

"Национальный исследовательский университет "Высшая школа экономики"

Факультет математики

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

На тему: "Марковские цепи и кривизна Риччи"

Студент группы N° 4.11.1
Рогов Андрей Анатольевич

Научный руководитель
доцент, доктор физ.-мат. наук
Колесников Александр Викторович

Москва, 2013 г.

Аннотация

В представленной работе мы рассматриваем понятие кривизны Риччи марковских цепей на метрических пространствах в контексте транспортной задачи. Вычисление кривизны Риччи в прикладных задачах, описываемых марковскими процессами, позволяет оценить скорость их сходимости к стационарному распределению. Нашей целью является исследование скорости сходимости в задаче кинетики социального неравенства с помощью кривизны Риччи.

Содержание

Введение

В ряде прикладных задач физики, информатики, экономики и других областей знаний, при построении математических моделей мы получаем марковские цепи. При этом важную роль играет оценка скорости сходимости марковского процесса к стационарному состоянию и его наличие. Одним из математических инструментов, позволяющих исследовать данный вопрос, является понятие кривизны Риччи. В частности, если кривизна Риччи марковской цепи между любыми двумя точками метрического пространства, на котором она определена, положительна, то мы имеем сжимаемость расстояния между потомками двух произвольных стартовых распределений марковского процесса на каждом последующем его шаге. Чем выше кривизна Риччи, тем выше скорость сходимости к стационарному распределению. В данной работе рассматривается понятие кривизны Риччи для ряда примеров и ее применение для исследования скорости сходимости прикладной экономической задачи кинетики социального неравенства. Показано, что в данной задаче скорость сходимости ограничена снизу (и сверху) для любых стартовых вероятностных распределений.

Ключевые слова: *цепи Маркова, кривизна Риччи, случайное блуждание, оптимальная стоимость транспортировки*

Глава 1. Понятие локальной и нелокальной кривизны Риччи для марковских цепей на метрических и конечных пространствах

§1 Локальная кривизна Риччи цепей Маркова на метрических пространствах

Для введения понятия локальной (между двумя точками x и y) кривизны Риччи, рассмотрим произвольное метрическое пространство X. На нем может быть задано некоторое вероятностное распределение ν, которое мы интерпретируем как единичную массу, распределенную по этому пространству. Ясно, что суммарная масса может быть произвольной, но далее она нормируется к единичной, что делает ее вероятностным распределением на X. Чтобы построить марковскую цепь на этом пространстве, мы вводим семейство вероятностных распределений $m_x(\cdot)$, определяющих вероятностные переходы из малой окрестности каждой точки x пространства X в малые окрестности всех точек $(\cdot)$ пространства. Каждый элемент такого семейства легко интерпретировать как план перераспределения (транспортировки) единичной массы из точки x во все точки X за один шаг марковской цепи (что соответствует частному случаю заданного стартового распределения, когда единичная масса оказывается сосредоточенной в точке x). Таким образом, мы понимаем семейство мер $m_x(\cdot)$ как заданный фиксированный план транспортировки, показывающий в долевом соотношении, куда транспортируется масса, оказавшаяся в малой окрестности точки . Совершив один шаг в марковской цепи, мы, с помощью этого семейства распределений, получим из исходного заданного распределения новое распределение. В частном случае конечного пространства X мы получаем обычную дискретную марковскую цепь, где семейство $m_x(\cdot)$ соответствует матрице переходов, а отдельный элемент семейства - строке этой матрицы. Мы можем многократно применять такой план транспортировки, совершая шаги марковской цепи и получая на каждом шаге новое распределение массы. Ясно, что при определенных видах плана перевозок (при условии эргодичности) на бесконечно удаленном шаге результативное распределение может сходиться к стационарному распределению, индиффирентному по отношению к заданному плану перевозок. В конечной марковской цепи это собственный вектор матрицы переходов с собственным числом равным единице. По аналогии с конечной цепью Маркова можно считать, что эргодичность обеспечивается наличием такого интегрального плана за конечное число шагов, который обеспечивает ненулевую перевозку из любой малой окрестности x в любую малую окрестность y(что соответствует существованию некоторой конечной степени матрицы переходов, не содержащей нулевых элементов).

Прежде всего мы хотим дать базовые определения, которые дает Yann Ollivier[2]:

Определение 1.1

Пусть (X, d) - польское метрическое пространство, снабженное борелевской σ-алгеброй. Случайное блуждание m на X есть семейство вероятностных мер $m_x(\cdot)$ на X для любого $x \in X$, удовлетворяющее двум следующим предположениям: (i) мера m_x зависит от точки $x \in X$; (ii) каждая мера m_x имеет конечный первый момент, т.е. для любого $o \in X$ и для любого

$x \in X$ имеем $\int d(o,y)dm_x(y) < \infty$.

Определение 1.2

Пусть (X,d) - метрическое пространство и пусть ν_1, ν_2 - две вероятностные меры на X. Вводится метрика: расстояние между ν_1 и ν_2 есть

$$\tau(\nu_1,\nu_2) := \inf_{\xi \in \Pi(\nu_1,\nu_2)} \int_{(x,y)\in X\times X} d(x,y)d\xi(x,y)$$

где $\Pi(\nu_1,\nu_2)$ - это множество мер на $X \times X$ проецируемое на ν_1 и ν_2.

Определение 1.3

Пусть (X,d) - метрическое пространство со случайным блужданием m. Пусть $x, y \in X$ - две различные точки. Локальная кривизна Риччи между парой точек x, y пространства (X,d,m) определяется формулой:

$$\kappa(x,y) := 1 - \frac{\tau(m_x,m_y)}{d(x,y)}$$

Мы хотим пояснить смысл $\kappa(x,y)$. Вероятностная мера ν на X может пониматься как единичная масса, распределенная на x (в дискретном случае это масса, распределенная по вершинам). Семейство вероятностных мер $m_x(\cdot)$ на X - это заданный план транспортировки для масс. Стоимость транспортировки есть произведение массы на тариф транспортировки. В качестве тарифа транспортировки мы рассматриваем расстояние, то есть $d(x,y)$ - это стоимость транспортировки единичной массы из x в y.

Распределение масс ν_1 может быть перенесено в распределение ν_2 путем переброски некоторых масс на определенные расстояния. Оптимальный способ транспортировки гарантирует нам минимальную стоимость.

Поэтому $\tau(\nu_1, \nu_2)$ - это оптимальная (минимальная) стоимость транспортировки и m_x есть результат переноса единичной массы из x в y согласно заданному плану транспортировки.

Кривизна Риччи $\kappa(x, y)$ дает нам возможность сравнить стоимость оптимальной транспортировки единичной массы из распределения m_x в распределение m_y (то есть $\tau(m_x, m_y)$) и стоимость транспортировки той же массы из x в y(то есть $d(x, y)$).

Если $\tau(m_x, m_y) < d(x, y)$, то $\kappa(x, y) > 0$;

если $\tau = d$, то $\kappa = 0$;

и если $\tau > d$, то $\kappa < 0$.

Определение 1.4

Мера ν на X называется *инвариантной* для случайного блуждания если $d\nu(x) = \int_y d\nu(y) dm_y(x)$.

Определение 1.5

Мера ν на X называется *реверсивной*, если она инвариантна и $d\nu(x)dm_x(y) = d\nu(y)dm_y(x)$.

Если мы рассматриваем дискретный случай, где X - конечное множество, состоящее из n элементов, тогда семейство вероятностных мер $m_x(\cdot)$ принимает вид матрицы перехода $\mathbb{K}$ с элементами $K(x, y) = m_x(y)$, $x, y \in X$, где $m_x(y)$ вероятность переноса массы из x в y. Это означает, что если масса $\nu(x)$ расположена в вершине x то, согласно плану транспортировки $\mathbb{K}$, масса, равная $\nu(x)m_x(y)$ будет перенесена из x в y.

В этом случае, инвариантное распределение $\nu = (\nu(x_1), \dots, \nu(x_n))$ является собственным вектором матрицы $\mathbb{K}$, т.е., $\nu = \nu\mathbb{K}$ или $\nu(x) = \sum_{y\in X} \nu(y)m_x(y)$. Распределение ν обратимо, если оно инвариантно и $\nu(x)m_x(y) = \nu(y)m_y(x)$, что означает, что для заданного распределения масс ν равные массы будут одновременно перенесены из x в y и из y в x в результате применения плана транспортировки $\mathbb{K}$.

Мы собираемся рассмотреть три примера из работы [2] и один пример, данный автором (Пример 1.3), и доказать их.

Пример 1.1 (Дискретный процесс Орнштейна-Уленбека)

Пусть $X = \{-N, -N+1, \dots, N, N+1\}$ и пусть m - случайное блуждание на X, заданное следующим образом: $m_k(k) = \frac{1}{2}, m_k(k+1) = \frac{1}{4} - \frac{k}{4N}, m_k(k-1) = \frac{1}{4} + \frac{k}{4N}$. Тогда биномиальное распределение $\frac{1}{2^{2N}}\binom{2N}{N+k}$ реверсивно, и для любых двух соседей x, y в X имеем $\kappa(x,y) = \frac{1}{2N}$

<u>Доказательство</u>:

Пусть $d(x,y) = l$ и пусть x и y соединены ребром длины l. Мы полагаем, что x - это k и y - это $k+1$. Представим вероятностные распределения m_x и m_y в виде следующей таблицы:

m_x	$\frac{1}{4}+\frac{k}{4N}$	$\frac{1}{2}$	$\frac{1}{4}-\frac{k}{4N}$	0
X	$k-1$	k	$k+1$	$k+2$
m_y	0	$\frac{1}{4}+\frac{k+1}{4N}$	$\frac{1}{2}$	$\frac{1}{4}-\frac{k+1}{4N}$

Очевидно, что оптимальный план транспортировки следующий:

- из вершины $k-1$ в вершину k мы переносим массу, равную $\frac{1}{4}+\frac{k}{4N}$;
- из k в $k+1$ мы переносим $\frac{1}{2}-\frac{1}{4N}$;
- из $k+1$ в $k+2$ мы переносим $\frac{1}{4}-\frac{k+1}{4N}$.

Расстояние транспортировки между m_x и m_y равно
$\tau(m_x, m_y) = l\left(\left(\frac{1}{4}+\frac{k}{4N}\right)+\left(\frac{1}{2}-\frac{1}{4N}\right)+\left(\frac{1}{4}-\frac{k+1}{4N}\right)\right) = l\left(1-\frac{1}{N}\right)$
Следовательно, $\kappa(x,y) = 1-\frac{\tau}{l} = \frac{1}{2N}$.

Проверим, что $\frac{1}{2^{2N}}\binom{2N}{N+k}$ - реверсивное распределение:

$$\frac{(2N)!}{(2N-N-k)!(N+k)!}\left(\frac{1}{4}-\frac{k}{4N}\right) = \frac{(2N)!}{(2N-N-k-1)!(N+k+1)!}\left(\frac{1}{4}+\frac{k}{4N}\right)$$

$$(N+k+1)\left(\frac{1}{4}-\frac{k}{4N}\right) = (2N-N-k)\left(\frac{1}{4}+\frac{k}{4N}\right)$$

$$N+k+1-k-\frac{k^2}{N}-\frac{k}{N} = N-k+k+1-\frac{k^2}{N}-\frac{k}{N}$$

$$N+1-\frac{k^2}{N}-\frac{k}{N} = N+1-\frac{k^2}{N}-\frac{k}{N}$$

Таким образом, $\frac{1}{2^{2N}}\binom{2N}{N+k}$ - реверсивное распределение для данного случайного блуждания.

Пример 1.2 (Мультиномиальное распределение)

Рассмотрим множество $X = (x_0, x_1, \ldots, x_d), x_i \in N, \sum x_i = N$ как набор из N шаров, размещенных в $d+1$ коробке. Рассмотрим процесс, заключающийся в вынимании из коробки одного из N шаров случайным образом и возвращении его случайным образом в одну из $d+1$ коробок. Более точно, вероятность переноса из $(x_0, \ldots, x_d)$ в $(x_0, \ldots, x_{i-1}, \ldots, x_{j+1}, \ldots, x_d)$ (возможен случай $i=j$) равна $\frac{x_i}{N(d+1)}$. Мультиномиальное распределение $\frac{N!}{(d+1)^N \Pi x_i!}$ является реверсивным для данной цепи Маркова. Снабдим конфигурационное пространство метрикой $d((x_i),(y_i)) \colon = \frac{1}{2}\sum |x_i - y_i|$, которая задает расстояние на графе. Тогда кривизна Риччи равна $\frac{1}{N}$.

Доказательство:

Обозначим $z_{ij} = (z_0, \dots, z_i - 1, \dots, z_j + 1, \dots, z_d), z \in X$. Рассмотрим пару вершин (x, y), где $y = x_{ij} = (x_0, \dots, x_i - 1, \dots, x_j + 1, \dots, x_d) = (y_0, \dots, y_i, \dots, y_j, \dots, y_d)$, $d(x, y) = 1$.

Вероятностное распределение m_x отправляет массу, равную 1, из вершины x в вершину $x_{pk}, p, k = 0, 1, \dots, d$. Распределение m_y - аналогично. Осуществим оптимальную транспортировку массы из m_x в m_y. Это есть транспортировка из x_{pk} в y_{pk} на расстояние, равное $d(x_{pk}, y_{pk}) = 1$, так как $x_{pk} = (x_0, \dots, x_p - 1, \dots, x_k + 1, \dots, x_d)$ и $y_{pk} = (x_0, \dots, x_p - 1, \dots, x_k + 1, \dots, x_i - 1, \dots, x_j + 1, \dots, x_d)$ (например, $i, j \neq p, k$).

Если $p \neq i, j$, то $m_x(x_{pk}) = \frac{x_p}{N(d+1)} = \frac{y_p}{N(d+1)} = m_y(y_{pk})$. Эти массы переносятся полностью.

Если $p = i$, то $m_x(x_{ik}) = \frac{x_i}{N(d+1)}; m_y(y_{ik}) = \frac{y_i}{N(d+1)} = \frac{x_i - 1}{N(d+1)}$. Таким образом, из всех вершин x_{ik} мы перенесем массу, равную $\frac{x_i - 1}{N(d+1)}$ и оставим массу, равную $\frac{1}{N(d+1)}$. (Суммарная масса по всем точкам равна $\frac{x_i - 1}{N}$) Тогда в вершинах x_{ik} будет суммарная масса, равная $\frac{1}{N}$ (эту массу следует перенести, но мы не знаем, куда).

Если $p = j$, то $m_x(x_{jk}) = \frac{x_j}{N(d+1)}$ и $m_y(y_{jk}) = \frac{y_j}{N(d+1)} = \frac{x_j + 1}{N(d+1)}$

Из всех вершин x_{jk} мы переносим суммарную массу, равную $\frac{x_j}{N}$, в вершины y_{jk}. Таким образом, нам нужно перенести добавочную массу $\frac{1}{N}$ в вершины y_{jk}. Тогда мы переносим массу, равную $\frac{1}{N}$, из x_{ik} в y_{jk}.

Заметим, что $x_{ik} = (x_0, \dots, x_i - 1, \dots, x_k + 1, \dots, x_j, \dots, x_d) = y_{jk}$ и $d(x_{ik}, y_{jk}) = 0$. Таким образом, масса, равная $\frac{1}{N}$, не переносится и

$$\tau(m_x, m_y) = \left(1 - \frac{1}{N}\right) d(x_{pk}, y_{pk}) = 1 - \frac{1}{N}$$

$$\kappa(x,y) = 1 - \frac{1-\frac{1}{N}}{d(x,y)} = \frac{1}{N}.$$

Покажем, что мультиномиальное распределение $\frac{N!}{(d+1)^N \Pi x_i!}$ реверсивно для этой марковской цепи.

$$\frac{N!}{(d+1)^N \Pi y_k!} \cdot \frac{y_j}{N(d+1)} = \frac{N!}{(d+1)^N (x_i-1)!(x_j+1)! \prod_{k \neq i,j} x_k!} \cdot \frac{(x_j+1)}{N(d+1)} =$$

$= \frac{N!}{(d+1)^N \prod x_k!} \cdot \frac{x_i}{N(d+1)}$.

Пример 1.3 (Дискретный N-мерный тор)

Рассмотрим дискретный N-мерный тор. Он представляет собой замкнутую цепь из конечного числа вершин (N_i, где $i = 1, \dots, n$) вдоль каждой координаты i. Пусть транспозиция δ_{+i} означает переход в последующую (соседнюю) вершину по координате i в направлении, выбранном как положительное (по часовой стрелке в цикле с координатой i), и транспозиция δ_{-i} означает переход в предыдущую (соседнюю) вершину по координате i в направлении, выбранном как отрицательное (против часовой стрелки в цикле с координатой i).Расстояние между соседними вершинами в X для определенности задаем равными единице ($d(x, \delta_{\pm i} x) = 1$). Введем на X семейство вероятностных мер $m_x(\cdot)$ следующим образом:

Пусть $x, y \in X$ соседи и $y = \delta_{+i} x$. Пусть для каждого $x \in X$,

$$m_x(\delta_{+j} x) = p_{+j}, m_x(\delta_{-j} x) = p_{-j}, m_x(x) = \lambda$$

$$m_y(\delta_{+j} y) = p_{+j}, m_y(\delta_{-j} y) = p_{-j}, m_y(y) = \lambda,$$

где $(p_{+j} + p_{-j} + \lambda) = 1$.

Кривизна Риччи для этой цепи Маркова равна $\kappa(x,y) = 0$.

Доказательство:

Если $i \neq j$, то из вершин $\delta_{\pm j}x$ мы перенесем массу, равную $p_{\pm j}$, в вершины $\delta_{\pm j}y$ на расстояние, равное $d(\delta_{\pm j}x, \delta_{\pm j}y) = 1$. В этом случае массы перенесены полностью.

Если $i = j$, то проверим, существует ли такая масса, которую не надо переносить. Для этого будет полезно воспользоваться следующей таблицей:

m_x	p_{-i}	λ	p_{+i}	0
X	$\delta_{-i}x$	x	$\delta_{+i}x = y$	$\delta_{+i}y$
m_y	0	p_{-i}	λ	p_{+i}

Из таблицы видно, что

- из вершины $\delta_{-i}x$ в вершину x мы переносим массу, равную p_{-i},
- из вершины y в вершину $\delta_{+i}y$ мы переносим p_{+i}
- из x в y мы переносим λ.

Следовательно, в этом случае все массы перенесены полностью.

Таким образом, $\tau(m_x, m_y) = \left(\sum_{j=1}^{n}(p_{-j} + p_{+j}) + \lambda\right) d(x,y) = d(x,y)$ и $\kappa(x,y) = 0$

Пример 1.4 (Дискретный n-мерный куб)

Рассмотрим n-мерный куб $X = Q^n$, где $x = (x_1, \dots, x_n), x_i \in \{0,1\}$. Положим $y = \delta_i x = (x_1, \dots, \bar{x}_i, \dots, x_n)$, где $\bar{}$ - операция отрицания. Обозначим через δ_i инверсию i-той координаты.

Пусть $K(x,x) = \lambda$; $K(x,z) = p_j$, где $z = \delta_j x$ для $j = 1, \dots, n$. Пусть

$d(x, z) = l_j$ и

$$m_x(\delta_j x) = p_j, \quad m_y(\delta_j y) = p_j, \quad m_y(y) = m_x(x) = \lambda, \quad m_y(x) = m_x(y) = p_i$$

Кривизна Риччи для этой цепи Маркова равна $\kappa(x, y) = 2\min(\lambda, p_i)$.

Доказательство:

Если $i \neq j$, то мы переносим массу, равную p_j из вершины $\delta_j x$ в вершину $\delta_j y$ на расстояние l_i.

Если $i = j$, то мы переносим массу, равную $|p_i - \lambda|$ из x в y. Это представляется в следующей таблице:

m_x	λ	p_i
X	x	y
m_y	p_i	λ

Тогда $\tau(x, y) = \left(\sum_{i \neq j} p_j + |p_i - \lambda|\right) l_i = (1 - p_i - \lambda + |p_i - \lambda|) l_i$
и $\kappa(x, y) = 1 - \frac{\tau}{l_i} = p_i + \lambda - |p_i - \lambda| = 2\min(\lambda, p_i)$.

§2 Нелокальная кривизна Риччи конечных цепей Маркова

Понятие нелокальной кривизны Риччи для конечных марковских цепей, заданных матрицей перехода, вводится с использованием принципиально иных подходов. Оно является единой интегральной характеристикой марковской цепи и строится на базе стационарного распределения π, ортогональной ему гиперплоскости ρ (проходящей через вектор из единичных компонент), и введенной на ней специальной $\mathcal{W}$-метрике. Сама нелокальная кривизна Риччи вводится как неравенство с ограничением, основанном на кривизне функции энтропии от аргумента из пространства ρ с $\mathcal{W}$-метрикой.

Последующие определения и идеи взяты из работы Маттиаса Эрбара и Яна Мааса[1].

Рассмотрим конечную марковскую цепь на множестве вершин X с матрицей переходов $\mathbb{K}$. Пусть π - инвариантная вероятностная мера на X $(\pi = \pi\mathbb{K})$.

Пусть

$$\mathcal{P}(X) := \left(\rho\colon X \to \mathbb{R}_+ | \quad \sum_{x\in X} \pi(x)\rho(x) = 1\right)$$

- множество плотностей вероятности на X. Подмножество, состоящее из этих плотностей вероятности строго положительно и обозначается как $\mathcal{P}_*(X)$. Мы рассматриваем метрику $\mathcal{W}$, определяемую для $\rho_0, \rho_1 \in \mathcal{P}(X)$ следующим образом:

$$\mathcal{W}(\rho_0, \rho_1)^2 := \inf_{\rho,\psi} \left(\frac{1}{2}\int_0^1 \sum_{x,y\in X} (\psi_t(x) - \psi_t(y))^2 \hat{\rho}_t(x,y)K(x,y)\pi(x)dt\right),$$

где инфимум пробегает по всем кривым $\rho\colon [0,1] \to \mathcal{P}(X)$ и $\psi\colon [0,1] \to \mathbb{R}^X$, удовлетворяющим уравнению непрерывности

$$\begin{cases} -\dfrac{d}{dt}\rho_t(x) + \sum_{y\in X}(\psi_t(x) - \psi_t(y))\hat{\rho}_t(x,y)K(x,y) = 0, & \forall x \in X \\ -\rho(0) = \rho_0, \qquad \rho(1) = \rho_1. \end{cases}$$

Где $\hat{\rho}(x,y) := \int_0^1 \rho(x)^{1-p}\rho(y)^p dp$. Важность значения логарифма в этом контексте связана с тождеством

$$\rho(x) - \rho(y) = \hat{\rho}(x,y)(\log\rho(x) - \log\rho(y)).$$

Так как метрика $\mathcal{W}$ является римановой в $\mathcal{P}_*(X)$, то имеет смысл рассматривать градиентные потоки в $(\mathcal{P}_*(X), \mathcal{W})$, и было доказано, что тепловой поток энтропии, связанный с непрерывной марковской полугруппой $P_t = e^{t(K-I)}$, есть градиентный поток энтропии

$$\mathcal{H}(p) = \sum_{x \in X} \pi(x)\rho(x)\log\rho(x)$$

по отношению к римановой структуре, определяемой $\mathcal{W}$.

Далее мы основываем наше доказательство на следующем определении и предложении из [1]:

Определение 2.1

Мы говорим, что $\mathbb{K}$ обладает нелокальной кривизной Риччи, ограниченной снизу числом $\kappa \in \mathbb{R}$, если для любой геодезической постоянной скорости $\{\rho_t\}_{t\in[0,1]}$ в $(\mathcal{P}_*(X), \mathcal{W})$ имеет место

$$\mathcal{H}(p) \leq (1-t)\mathcal{H}(\rho_0) + t\mathcal{H}(\rho_1) - \frac{\kappa}{2}t(1-t)\mathcal{W}(\rho_0, \rho_1)^2$$

Вышеприведенные формулы мотивированы известной формулой Бенаму-Бренье. В непрерывном случае известно, что аналогичный подход приводит к определению кривизны, эквивалентному классическому.

Каждая неприводимая, реверсивная марковская цепь обладает некоторым множеством отображений (переходов). На самом деле, явный вид множества отображений может быть описан следующим образом: для $x, y \in X$ рассмотрим биекцию $t_{\{x,y\}} \colon X \to X$, которая меняет местами x и y и сохраняет остальные точки фиксированными. Тогда пусть G есть множество всех таких транспозиций, и рассмотрим набор весов $c(x, t_{\{x,y\}}) = K(x, y)$ и $c(x, t_{\{y,z\}}) = 0$ для $x \notin \{y, z\}$. Тогда (G, c) определяет множество отображений.

Предложение 2.1

Пусть K - матрица переходов неприводимой и реверсивной марковской цепи на конечном множестве X, и пусть (G, c) - множество отображений для нее. Рассмотрим следующие условия:

$i) \delta \circ \eta = \eta \circ \delta, \quad$ для всех $\delta, \eta \in G$,

$ii) c(\delta x, \eta) = c(c, \eta)$, $\quad$ для всех $x \in X, \delta, \eta \in G$,

$iii) \delta \circ \delta = id, \quad$ для всех $\delta \in G$.

Если выполнены условия $i)$ и $ii)$, тогда Ric$(K) \geq 0$. Если дополнительно выполняется $iii)$, тогда Ric$(K) \geq 2C$, где

$$C := \min\{c(x, \delta) \colon x \in X, \delta \in G \quad \text{, такое что} \quad c(x, \delta) > 0\}$$

Пример 2.1 (Дискретный N-мерный тор)

Рассмотрим дискретный N-мерный тор. Он представляет собой замкнутую цепь из конечного числа вершин (N_i, где $i = 1, \dots, n$) вдоль каждой координаты i. Пусть транспозиция δ_{+i} означает переход в последующую (соседнюю) вершину по координате i в направлении, выбранном как положительное (по часовой стрелке в цикле с координатой i), и транспозиция δ_{-i} означает переход в предыдущую (соседнюю) вершину по координате i в направлении, выбранном как отрицательное (против часовой стрелки в цикле с координатой i).Расстояние между соседними вершинами в X для определенности задаем равными единице ($d(x, \delta_{\pm i} x) = 1$). Введем на X семейство вероятностных мер $m_x(\cdot)$ следующим образом:

Пусть $x, y \in X$ соседи и $y = \delta_{+i} x$. Пусть для каждого $x \in X$,

$$m_x(\delta_{+j} x) = p_{+j}, m_x(\delta_{-j} x) = p_{-j}, m_x(x) = \lambda$$

$$m_y(\delta_{+j}y) = p_{+j}, m_y(\delta_{-j}y) = p_{-j}, m_y(y) = \lambda,$$

где $(p_{+j} + p_{-j} + \lambda) = 1$.

Тогда нелокальная кривизна Риччи для этой цепи Маркова ограничена снизу нулем.

Доказательство:

$G = \{+i, -i, id\}(i = 1, \dots, n)$

$c(x, \delta_{-i}) = p_{-i}, c(x, \delta_{+i}) = p_{+i}$ и $c(x, id) = \lambda$.

Тогда пункты $i)$ и $ii)$ Предложения 2.1 удовлетворены и $Ric(K) \geq 0$

Пример 2.2 (Дискретный n-мерный куб)

Рассмотрим n-мерный куб $X = Q^n$, где $x = (x_1, \dots, x_n), x_i \in \{0, 1\}$. Положим $y = \delta_i x = (x_1, \dots, \bar{x}_i, \dots, x_n)$, где $\bar{}$ - операция отрицания. Обозначим через δ_i инверсию i-той координаты.

Пусть $K(x, x) = \lambda$; $K(x, z) = p_j$, где $z = \delta_j x$ для $j = 1, \dots, n$. Пусть $d(x, z) = l_j$ и

$$m_x(\delta_j x) = p_j, \quad m_y(\delta_j y) = p_j, \quad m_y(y) = m_x(x) = \lambda, \quad m_y(x) = m_x(y) = p_i$$

Тогда нелокальная кривизна Риччи для этой цепи Маркова ограничена снизу числом $2\min(p_1, \dots, p_n, \lambda)$.

Доказательство:

Let $G = \{\delta_1, \dots, \delta_n, id\}$,

$c(x, \delta_i) = p_i$ и $c(x, id) = \lambda, \quad (i = 1, \dots, n)$.

Тогда пункты $i)$, $ii)$ и $iii)$ Предложения 2.1 удовлетворены и $Ric(K) \geq 2\min(p_1, \dots, p_n, \lambda)$.

Понятие локальной кривизны Риччи для марковской цепи на метрическом пространстве имеет ясную интерпретацию с точки зрения транспортной задачи. Она характеризует свойства марковской цепи локально между двумя точками x и y. Для большого числа примеров локальная кривизна Риччи может быть найдена непосредственно с использованием ее определения. Причем наличие метрики играет базовую роль, так как кривизна зависит от расстояний транспортировки по различным маршрутам. Понятие нелокальной кривизны Риччи характеризует свойства конечной марковской цепи в целом, базируясь на стационарном распределении. Она совершенно не учитывает длину ребер графа, соответствующего марковской цепи и имеет принципиально иную схему построения, чем локальная кривизна Риччи. Имеет место существенная сложность конструирования нелокальной кривизны Риччи в связи со сложностью вычисления расстояния в используемой $\mathcal{W}$-метрике. Непосредственно по определению нелокальную кривизну Риччи можно легко вычислить лишь для самых простейших задач. Для этих простейших задач наблюдается соответствие полученных значений локальной и нелокальной кривизн Риччи в случае, если длины ребер графа равны и тем самым исчезает влияние фактора расстояния транспортировки. Для родственных задач, где длины ребер графа существенно различны, могут быть получены совершенно разные значения локальной кривизны Риччи при неизменной нелокальной кривизне. Понятие нелокальной кривизны Риччи также можно распространить на марковские цепи метрических пространств, заменив суммирование интегрированием в основных определениях. В итоге отметим, что понятия локальной и нелокальной кривизн Риччи являются особыми характеристиками марковских цепей, отражающими их индивидуальные свойства.

Глава 2. Применение кривизны Риччи к определению скорости сходимости в задаче кинетики социального неравенства.

§1 Верхняя и нижняя оценки скорости сходимости задачи кинетики социального неравенства

Рассмотрим задачу кинетики социального неравенства.

Состояниями марковской цепи являются векторы $x = (x_1, \dots, x_N)$, где x_i - количество монет у жителя i, P - матрица переходов, $\nu = (\nu_1, \dots, \nu_x, \dots)$ - вероятностное распределение на множестве состояний X. Введем обозначение $P\nu$, где P - оператор, который означает результат матричного произведения $\nu \cdot P$ ($P^n\nu$ - соответственно $\nu \cdot P^n$). Обозначим $\delta_x = (0, \dots, 0, \nu_x = 1, 0, \dots)$. Обозначим $x_{ij} = (x_1, \dots, x_i + 1, \dots, x_j - 1, \dots, x_N)$, $i \neq j$. Зафиксируем произвольного соседа $y = x_{rs} = (\dots, x_r + 1, \dots, x_s - 1, \dots)$.

<u>Лемма 0.</u>

1).Пусть $x_r \geq 1, x_s \geq 2$. Тогда $W(P\delta_x, P\delta_y) = W(\delta_x, \delta_y) = d(x, y) = 1$

2).Пусть $x_r = 0, x_s \geq 2$. Тогда $W(P\delta_x, P\delta_y) = (1 - \frac{1}{N(N-1)})W(\delta_x, \delta_y)$

3).Пусть $x_r \geq 1, x_s = 1$. Тогда $W(P\delta_x, P\delta_y) = (1 - \frac{1}{N(N-1)})W(\delta_x, \delta_y)$

4).Пусть $x_r = 0, x_s = 1$. Тогда $W(P\delta_x, P\delta_y) = (1 - \frac{1}{N})W(\delta_x, \delta_y)$.

5).Пусть $x_r \geq 1, x_s \geq 2$, $x_r + 1 \leq x_s$. Обозначим $m = x_r + 1$. Тогда
$W(P^i\delta_x, P^i\delta_y) = W(\delta_x, \delta_y), i = 1, \dots, m-1;$
$W(P^m\delta_x, P^m\delta_y) = \left(1 - \frac{1}{N^m(N-1)}\right) W(\delta_x, \delta_y)$

6).Пусть $x_r \geq 1, x_s \geq 2$, $x_s \leq x_r + 1$. Обозначим $m = x_s$. Тогда
$W(P^i\delta_x, P^i\delta_y) = W(\delta_x, \delta_y), i = 1, \dots, m-1;$
$W(P^m\delta_x, P^m\delta_y) = \left(1 - \frac{1}{N^m(N-1)}\right) W(\delta_x, \delta_y)$.

<u>Доказательство:.</u>

Пока предположим, что $x_i \geq 1, \quad \forall i \neq r, s$:

1). $P\delta_x$ имеет следующее распределение весов по вершинам:

в вершине x_{ij} вес $\frac{1}{N(N-1)} \forall i \neq j$,

в вершине y_{ij} вес $\frac{1}{N(N-1)} \forall i \neq j$.

Осуществляется перенос весов $\frac{1}{N(N-1)}$ между вершинами x_{ij} и y_{ij}. В сумме переносится вес 1 на расстояние $d(x_{ij}, y_{ij}) = 1$

2). $P\delta_x$: вершин x_{ir} не существует ($\forall i \neq r$) и веса, которые должны были бы в них находится, остаются в x. То есть в вершине x вес $\frac{1}{N(N-1)} \cdot (N-1)$, в вершинах $x_{ij}, j \neq r$, веса $\frac{1}{N(N-1)}$.

$P\delta_y$: в вершинах y_{ij} веса $\frac{1}{N(N-1)}$, в том числе в вершинах y_{ir}, среди которых $y_{sr} = x$. Тогда для $j \neq r$ переносим веса $\frac{1}{N(N-1)}$ из x_{ij} в y_{ij} (на расстояние 1). Заметим, что т.к. $y_{ir} = (x_1, \ldots, x_r, \ldots, x_i + 1, \ldots, x_s - 1, \ldots)$ для $i \neq s$, то $d(x, y_{ir}) = 1$. Переносим $(N-2)$ раза веса $\frac{1}{N(N-1)}$ из x в вершины y_{ir}, $i \neq s$. Так как $d(x, y_{sr}) = d(x, x) = 0$, то вес $\frac{1}{N(N-1)}$ из x в y_{sr} не переносится. Тогда из веса 1 все переносится на расстояние 1, кроме веса $\frac{1}{N(N-1)}$

3). $P\delta_x$: в вершинах x_{ij} веса $\frac{1}{N(N-1)}$, в вершине x вес 0.

$P\delta_y$: вершин y_{is} не существует, а веса $\frac{1}{N(N-1)}$, которые должны были бы в них быть, остаются в y, то есть в y вес $\frac{1}{N(N-1)} \cdot (N-1)$. В вершинах $y_{ij}, j \neq s$, веса $\frac{1}{N(N-1)}$, в том числе в вершине $y_{sr} = x$. Заметим, что т.к. $x_{is} = (\ldots, x_r, \ldots, x_i + 1, \ldots, x_s - 1, \ldots)$, $y = (\ldots, x_r + 1, \ldots, x_i, \ldots, x_s - 1, \ldots)$, то для $i \neq r$ $d(x_{is}, y) = 1$; $d(x_{rs}, y) = d(y, y) = 0$. Тогда из $x_{ij}, j \neq s$, переносим веса $\frac{1}{N(N-1)}$ в y_{ij}. Из x_{is} переносим веса $\frac{1}{N(N-1)}$ в y $(N-2)$ раза. Из x_{rs} не переносится $\frac{1}{N(N-1)}$ в y. Т.е. вес 1 переносится на расстояние 1 за исключением веса $\frac{1}{N(N-1)}$.

4). $P\delta_x$: вершин x_{ir} не существует, в вершине x находится вес $\frac{1}{N(N-1)} \cdot$

$(N-1)$. В вершинах $x_{ij}, j \neq r$, находятся веса $\frac{1}{N(N-1)}$, в том числе в вершине $x_{rs} = y$.

$P\delta_y$: вершин y_{is} не существует, в вершине x находится вес $\frac{1}{N(N-1)} \cdot (N-1)$. В вершинах $y_{ij}, j \neq s$, находятся веса $\frac{1}{N(N-1)}$, в том числе в вершине $y_{sr} = x$. Заметим, что $x_{is} = (\ldots, x_r, \ldots, x_i + 1, \ldots, x_s - 1, \ldots)$, $y_{ir} = (\ldots, x_r, \ldots, x_i + 1, \ldots, x_s - 1, \ldots)$. То есть для $i \neq s, r$, $d(x_{is}, y_{ir}) = 0$, эти вершины совпадают. Но в них находятся веса $\frac{1}{N(N-1)}$. Мы их условно переносим из x_{is} в y_{ir} $(i \neq s, r)$ на расстояние 0. То есть не переносим вес $\frac{1}{N(N-1)} \cdot (N-2)$. Из $x_{ij}, j \neq s, r$, переносим веса $\frac{1}{N(N-1)}$ в y_{ij}. Из $x_{rs} = y$ в $y_{sr} = x$ не переносится $\frac{1}{N(N-1)}$. То есть в сумме не переносится вес $\frac{1}{N(N-1)} \cdot (N-2) + \frac{1}{N(N-1)} = \frac{1}{N}$.

Если некоторые компоненты $x_p, p \neq s, r$ вектора x равны нулю, то это не изменяет полученного результата, так как x_{ip}, y_{ip} будут отсутствовать, и их веса $\frac{1}{N(N-1)}$ останутся соответственно в x, y и будут перенесены из x в y на расстояние 1.

5). $P^i\delta_x; i = 1, \ldots, m-1$: на первом шаге будет переносится вес $\frac{1}{N(N-1)}$ в вершины x_{ir} (их $N-1$), что в сумме даст вес $\frac{1}{N}$. Так на каждом шаге. Мы отслеживаем переходы в те соседние вершины, при которых компонента r убывает на единицу на каждом шаге, при этом на шаге i компонента r всех рассматриваемых нами потомков становится равной $x_r - i$, а суммарная масса, перенесенная в них будет равна $\frac{1}{N^i}$. Именно эти отслеживаемые нами потомки самыми первыми достигнут нулевого значения компоненты r, причем одновременно, и тогда станет применим пункт 2) Леммы 0. У всех остальных потомков значение компоненты r будет больше и они позже, чем на шаге $m-1$, достигнут нулевого значения на компоненте r, поэтому они нами и не рассматриваются. Следует отметить, что пока $i \leq m-1$, значение компоненты r всех вообще потомков i-го поколения, перед шагом i, больше либо равно единице, а значение компоненты s больше либо равно двум. Тогда перед шагами $i = 1, \ldots, m-1$ применим пункт 1) Леммы 0 и для этих i мы получим равенство $W(P^i\delta_x, P^i\delta_y) = W(\delta_x, \delta_y)$.

В результате на шаге $m-1$ появятся вершины с нулевым значением компоненты r, которые будут обладать суммарным весом $\frac{1}{N^{m-1}}$. Тогда на шаге m из этих вершин не будет перенесен к соответствующим потомкам y вес $\frac{1}{N^{m-1}} \cdot \frac{1}{N(N-1)}$. Получим

$$W(P^m\delta_x, P^m\delta_y) = 1 - \frac{1}{N^m(N-1)} = \left(1 - \frac{1}{N^m(N-1)}\right) W(\delta_x, \delta_y)$$

6). $P^i\delta_x; i = 1, \ldots, m-1$: Аналогично пункту 5), мы отслеживаем переходы в те соседние вершины, при которых компонента s убывает на единицу на каждом шаге, при этом на шаге i компонента s всех рассматриваемых нами потомков становится равной $x_s - i$, а суммарная масса, перенесенная в них будет равна $\frac{1}{N^i}$. Именно эти отслеживаемые нами потомки самыми первыми достигнут единичного значения компоненты s, причем одновременно, и тогда станет применим пункт 3) Леммы 0. У всех остальных потомков значение компоненты s будет больше и они позже, чем на шаге $m-1$, достигнут единичного значения на компоненте s, поэтому не рассматриваются. Пока $i \leq m-1$ значение компоненты s всех вообще потомков i-го поколения, перед шагом i, больше либо равно двум, а $x_r \geq x_s - 1$. Тогда на шагах $i = 1, \ldots, m-1$ применим пункт 1) Леммы 0 и для этих i мы получим равенство $W(P^i\delta_x, P^i\delta_y) = W(\delta_x, \delta_y)$. Аналогично, для $P^{m-1}\delta_x$ мы получим на $(m-1)$ шаге потомков вершины x с компонентами r, равными 1, обладающими суммарным весом $\frac{1}{N^{m-1}}$. Тогда на шаге m соответствующим потомкам y не будет перенесена суммарная масса $\frac{1}{N^m(N-1)}$. Получим

$$W(P^m\delta_x, P^m\delta_y) = \left(1 - \frac{1}{N^m(N-1)}\right) W(\delta_x, \delta_y)$$

.

Доказано.

Следствие 1.

Таким образом, для любых соседей x, y ($d(x,y) = 1$), обозначив $m = min(x_r + 1, x_s)$, получим

$$W(P^m\delta_x, P^m\delta_y) = \left(1 - \frac{1}{N^m(N-1)}\right) W(\delta_x, \delta_y).$$

Замечание. Для $x_s \geq 1$ ситуацию $x_s = x_r + 1$ не рассматриваем, так как для того, чтобы перевести $(\ldots, x_r, \ldots, x_s, \ldots)$ в $(\ldots, 0, \ldots, 1, \ldots)$ потребуется $2(m-1)$ шагов, что невыгодно для дальнейших оценок.

Следствие 2.

Тогда $\forall \nu, \mu$ верна оценка

$$W(P^{n_0}\nu, P^{n_0}\mu) \leq \left(1 - \frac{1}{N^{n_0}(N-1)}\right) W(\nu, \mu),$$

где $n_0 = \frac{M}{2}$, если M - четное и $n_0 = [\frac{M}{2}] + 1$, если M - нечетное.

Доказательство:.

Так как при количестве монет M значение $m = min(x_r + 1, x_s)$ не может превышать $n_0 = \frac{M}{2}$, если M - четное, и $n_0 = [\frac{M}{2}] + 1$, если M - нечетное, то для произвольных соседей x, y верна общая оценка

$$W(P^{n_0}\delta_x, P^{n_0}\delta_y) \leq \left(1 - \frac{1}{N^{n_0}(N-1)}\right) W(\delta_x, \delta_y), \quad d(x,y) = 1$$

Следуя *Ollivier*, так как в данной задаче для любых двух вершин x, y : $d(x,y) = l$ можно перейти из x в y, совершив l одношаговых переходов $x = x^0 \to x^1 \to \ldots \to x^l = y$, то

$$W(P^{n_0}\delta_x, P^{n_0}\delta_y) \le \sum_{i=1}^{l} W(P^{n_0}\delta_{x^{i-1}}, P^{n_0}\delta_{x^i}) \le \left(1 - \frac{1}{N^{n_0}(N-1)}\right) l =$$

$= \left(1 - \frac{1}{N^{n_0}(N-1)}\right) W(\delta_x, \delta_y).$

Итак, $\forall x, y \in X$ верна оценка

$$W(P^{n_0}\delta_x, P^{n_0}\delta_y) \le \left(1 - \frac{1}{N^{n_0}(N-1)}\right) W(\delta_x, \delta_y)$$

(По *Ollivier*) Тогда $\forall \nu, \mu \Rightarrow$

$$W(P^{n_0}\nu, P^{n_0}\mu) \le \left(1 - \frac{1}{N^{n_0}(N-1)}\right) W(\nu, \mu),$$

где $n_0 = \frac{M}{2}$, если M - четное и $n_0 = [\frac{M}{2}] + 1$, если M - нечетное.

Доказано.

Везде в дальнейшем для простоты будем считать M четным.

Укажем явно пары состояний x, y, дающие минимальную и максимальную кривизны:

1)

$$x = (\ldots, 0, \frac{M}{2}, 0, \ldots, 0, \frac{M}{2}, 0, \ldots)$$
$$y = (\ldots, 0, \frac{M}{2} + 1, 0, \ldots, 0, \frac{M}{2} - 1, 0, \ldots)$$

дают

$$W(P^{\frac{M}{2}}\delta_x, P^{\frac{M}{2}}\delta_y) = \left(1 - \frac{1}{N^{\frac{M}{2}}(N-1)}\right) W(\delta_x, \delta_y)$$

- самая низкая (исключая нулевую для единичного шага)

2)

$$x = (\dots, 0, \dots, 1, \dots)$$
$$y = (\dots, 1, \dots, 0, \dots)$$

дают
$$W(P\delta_x, P\delta_y) = \left(1 - \frac{1}{N}\right) W(\delta_x, \delta_y)$$
- самая высокая (верхняя оценка скорости сходимости).

Рассмотрим задачи (N, M), для которых матрица переходов обратима (в их числе, например, два достаточно тривиальных представителя $(N = 2, M = 2), (N = 3, M = 1)$, для которых имеется всего три состояния, и соответствующие матрицы переходов $P_{3\times 3}$ обратимы). Отметим при этом , что P^{-1} будет содержать отрицательные элементы и что решая задачу $\mu P = \nu$, где ν - заданное вероятностное распределение, а μ - неизвестный вектор-строка, мы получим решение $\mu = \nu P^{-1}$, но μ может при этом содержать отрицательные компоненты, то есть μ не обязательно будет вероятностным распределением.

Докажем, что полученная нижняя оценка скорости сходимости неулучшаема (по крайней мере локально) для задач (N, M), матрица переходов которых обратима.

Утверждение.

Пусть (N, M) задача имеет обратимую матрицу переходов. Тогда
$\forall \nu^0 \quad (\nu^0_x > 0 \quad \forall x \in X), \forall n \quad \exists \mu^0$ такое, что:
$$W(P^{n+1}\nu^0, P^{n+1}\mu^0) = W(P^n\nu^0, P^n\mu^0);$$
$$\dots$$
$$W(P^{n+m-1}\nu^0, P^{n+m-1}\mu^0) = W(P^n\nu^0, P^n\mu^0);$$

$$W(P^{n+m}\nu^0, P^{n+m}\mu^0) = \left(1 - \frac{1}{N^m(N-1)}\right) W(P^n\nu^0, P^n\mu^0),$$

где $m = \frac{M}{2}$. При этом M, не ограничивая общности, для простоты считаем четным.

__Доказательство:__

Возьмем любые $n, \nu^0 > 0$. Возьмем $x = (\ldots, 0, \frac{M}{2}, 0, \ldots, 0, \frac{M}{2}, 0, \ldots)$, $y = (\ldots, 0, \frac{M}{2} + 1, 0, \ldots, 0, \frac{M}{2} - 1, 0, \ldots)$. Вершины x, y выбраны специальным образом, чтобы соответствовать пункту 6) Леммы 0. Именно этот выбор обеспечит равенства утверждения, так как мы создаем ситуацию, когда на первом (после n) шаге марковской цепи оптимальный перенос массы осуществляется исключительно между вершинами x, y, что впоследствии дает самую малую положительную кривизну, причем только на шаге $m = \frac{M}{2}$. Рассмотрим $\nu = P^n\nu^0$.

Пусть $\mu = \nu + q\delta x - q\delta y$, где q - некоторая положительная величина, которую мы можем произвольно выбрать сколь угодно малой, что не отразится на дальнейших рассуждениях. Необходимость возможности выбрать q сколь угодно малой обусловлена использованием в дальнейшем предельного перехода при $q \to 0$. На q дополнительно требуется наложить ограничения: $\nu_x + q \leq 1, \nu_y - q \geq 0$ - это необходимо для корректности последующих равенств. Выполнение этих равенств также обеспечивается возможностью выбора достаточно малого q (учитывая, что по условию $\nu^0 > 0$, тогда, так как матрица P не содержит единичных компонент, имеем $\nu > 0$, что обеспечивает $\nu_x \in (0, 1)$ для всех x). μ совпадает с ν кроме двух компонент $\mu_x = \nu_x + q, \mu_y = \nu_y - q$

Очевидно, что оптимальный план транспортировки из ν в μ - осуществить переброску массы q из вершины y в вершину x на расстояние $d(x, y)$

(вычитая q из компоненты ν_x и прибавляя к компоненте ν_y, получим из вектора ν вектор μ). Стоимость этой транспортировки в данном случае и составляет расстояние между векторами ν, μ, так как эта транспортировка единственна и равна $qd(x,y)$. Введем $\lambda = \nu - q\delta y$. Тогда $\nu = \lambda + q\delta y, \quad \mu = \lambda + q\delta x$. Компоненты вектора λ при оптимальной транспортировке не переносятся.

Поясним это подробнее. Компоненты векторов ν, μ, λ, соответствующие вершинам, отличным от x, y, равны между собой и переносы массы по ним не производятся. Компонента $\nu_x = \lambda_x = \mu_x - q$. Тогда, чтобы из ν_x получить μ_x, надо, сохранив и никуда не перемещая ν_x (а это и есть λ_x), добавить к ней q (взятое из ν_y). Компонента $\nu_y = \mu_y + q = \lambda_y + q, (\mu_y = \lambda_y)$. Тогда, чтобы из ν_y получить μ_y, надо забрать из ν_y массу q(перенести ее в ν_x), сохранив в составе ν_y массу λ_y (и никуда не перемещая λ_y). Таким образом, компоненты вектора λ, которые можно считать составной частью компонент вектора ν (равно как и вектора μ - это их общая часть), сохраняются неизменными, не участвуя в плане оптимального переноса. По-другому говоря, λ есть максимальная общая часть векторов ν и μ, $\lambda = (\ldots, \lambda_i = \min(\nu_i, \mu_i), \ldots)$, и ее компоненты (в составе вектора ν) должны быть оставлены на своих местах при трансформировании ν в μ (с помощью переброски масс между компонентами вектора ν), иначе транспортировка будет заведомо неоптимальной.

Следует также отметить тот факт, что λ не является вероятностным распределением, но компоненты λ неотрицательны. Чтобы трансформировать ν в μ, нельзя обойтись переносом меньшей массы чем q и перенести ее на меньшее расстояние, чем единичное, и, в то же время, такого переноса достаточно. Поэтому такой перенос является оптимальной транспортировкой и расстояние $W(\nu,\mu)$ не может быть меньше q, а в точности ему равно. При этом то, что $W(\nu,\mu)$ также может стать сколь угодно малым вследствие выбора q, не влияет на дальнейшие рассуждения.

Для иллюстрации приведем следующий пример:

Пусть $\nu = (0.5, 0.1, 0.4), \mu = (0.75, 0.1, 0.15), \nu_x = 0.5, \nu_y = 0.4, x, y$ - соседи. Тогда $\lambda = (0.5, 0.1, 0.15)$ - это общая часть ν, μ и эти компоненты не нужно подвергать транспортировке, они остаются в составе ν неизменными (иначе нарушится оптимальность), $q = 0.25$. Единственной возможной оптимальной транспортировкой становится перенос массы $q = 0.25$ из $\nu_y = 0.4$ в $\nu_x = 0.5$ на единичное расстояние, тогда мы получим требуемое $\mu = (0.75, 0.1, 0.15)$ и при этом $W(\nu,\mu) = q = 0.25$, уменьшить эту стоимость невозможно. Именно в таком смысле мы применяем фразу "компоненты вектора λ при оптимальной транспортировке не переносятся".

Причем ν, μ таковы, что обязательно произойдет перенос именно из y в x, причем единственный в составе оптимальной транспортировки, обеспечивающей $W(\nu,\mu)$. Кроме того отметим, что мы по ν, выбирая достаточно малое q (поясним, что уменьшая q, мы добиваемся гарантий того, что $\mu^0 = P^{-n}\mu > 0$, то есть μ^0 будет являться вероятностным распределением, о чем будет подробнее сказано ниже), получаем вектор μ, который, тем самым, является зависимым от выбора ν, q.

Получим

$$W(\nu,\mu) = qd(x,y) = qW(\delta x, \delta y) = q$$

$$P\nu = P\lambda + qP\delta y, P\mu = P\lambda + qP\delta x$$

$$P^2\nu = P^2\lambda + qP^2\delta y, P^2\mu = P^2\lambda + qP^2\delta x$$

$$\cdots\cdots\cdots$$

$$P^m\nu = P^m\lambda + qP^m\delta y, P^m\mu = P^m\lambda + qP^m\delta x$$

Компоненты векторов $P\lambda, P^2\lambda, \ldots, P^m\lambda$ также не переносятся при оптимальной транспортировке между потомками ν, μ. Это имеет место, так как раз λ является максимальной общей частью ν, μ, то и $P\lambda$ будет находиться в составе максимальной общей части векторов $P\nu, P\mu$, также $P^2\lambda, \ldots, P^m\lambda$ будут находиться в составе максимальной общей части пар векторов $P^2\nu, P^2\mu; \ldots; P^m\nu, P^m\mu$ соответственно. Тогда

$$W(P\nu, P\mu) = qW(P\delta x, P\delta y) = [\text{по пункту 6) Леммы 0}] = qW(\delta x, \delta y) = W(\nu, \mu)$$

$$\ldots\ldots\ldots$$

$$W(P^{m-1}\nu, P^{m-1}\mu) = qW(P^{m-1}\delta x, P^{m-1}\delta y) = [\text{по пункту 6) Леммы 0}] =$$
$$= qW(\delta x, \delta y) = W(\nu, \mu)$$
$$W(P^m\nu, P^m\mu) = qW(P^m\delta x, P^m\delta y) = q\left(1 - \frac{1}{N^m(N-1)}\right)W(\delta x, \delta y) =$$
$$= \left(1 - \frac{1}{N^m(N-1)}\right)W(\nu, \mu), \quad \left(m = \frac{M}{2}\right)$$

Искомое $\mu^0 = P^{-n}\mu = P^{-n}\nu + qP^{-n}\delta x - qP^{-n}\delta y = \nu^0 + qP^{-n}\delta x - qP^{-n}\delta y$, где P^{-n} - оператор, обратный P, примененный n раз.

Требуется позаботиться о том, чтобы полученный таким образом вектор μ^0 являлся бы также вероятностным распределением.

Так как $\lim_{q\to 0}\mu_0 = \lim_{q\to 0} P^{-n}\mu = P^{-n}\lim_{q\to 0}\mu = P^{-n}\nu = \nu_0 \Rightarrow \exists q\colon \mu^0 > 0$. То есть μ^0 - вероятностное распределение.

Тогда, учитывая, что $\nu = P^n\nu^0, \mu = P^n\mu^0$, будет иметь место

$$W(P^{n+i}\nu^0, P^{n+i}\mu^0) = W(P^n\nu^0, P^n\mu^0), i = 1, \ldots, m-1,$$
$$W(P^{n+m}\nu^0, P^{n+m}\mu^0) = \left(1 - \frac{1}{N^m(N-1)}\right)W(P^n\nu^0, P^n\mu^0)$$

Доказано.

Замечание. То же самое верно для всех других точных значений кривизны (за m шагов) между соседними состояниями. Данные кривизны могут реализоваться на любом сколь угодно большом шаге n для специально подобранных стартовых значений ν^0, μ^0. В том числе и для самой высокой кривизны: $\exists \nu^0, \mu^0 \colon \forall n$:

$$W(P^{n+1}\nu^0, P^{n+1}\mu^0) = \left(1 - \frac{1}{N}\right) W(P^n\nu^0, P^n\mu^0).$$

§2 Инструменты вычисления скорости сходимости конечных цепей Маркова с помощью кривизны Риччи.

Обозначим через b_{xy} массы, переносимые при оптимальном плане переноса из ν в μ между вершинами x и y (из вершин x в вершины y).

Лемма 1.

Пусть $W(\nu, \mu)$ реализуется оптимальной транспортировкой, при которой из вершин x в вершины y переносятся веса $b_{xy}, x \neq y$. Тогда имеет место неравенство:

$$W(P\nu, P\mu) \leq \left(1 - \sum_{x,y} \beta_{xy} k_{xy}\right) W(\nu, \mu),$$

где $\beta_{xy} = \frac{b_{xy} d(x,y)}{\sum_{x,y} b_{xy} d(x,y)}$.

Доказательство:

Рассмотрим $\forall \nu, \mu$. Обозначим $\lambda = (\ldots, \lambda_i = \min(\nu_i, \mu_i), \ldots)$. Введем следующие обозначения: $\Delta\nu = (\ldots, \nu_i - \min(\nu_i, \mu_i), \ldots)$, $\Delta\mu = (\ldots, \mu_i - \min(\nu_i, \mu_i), \ldots)$. То есть компонента λ_i - максимальная общая часть ν_i, μ_i и, очевидно, ее не требуется транспортировать при оптимальном плане переноса.

$$\nu = \lambda + \Delta\nu; \quad \mu = \lambda + \Delta\mu.$$

Массы $\Delta\nu_i$ полностью переносятся при оптимальной транспортировке из ν в μ

$$W(\nu,\mu) = W(\Delta\nu,\Delta\mu)$$

Рассмотрим

$$P\nu = P\lambda + P\Delta\nu; \quad P\mu = P\lambda + P\Delta\mu$$

$P\lambda$ также является составляющей непереносимой части при оптимальном переносе из $P\nu$ в $P\mu$. Однако, если кривизна Риччи положительна, то между $P\Delta\nu$ и $P\Delta\mu$ не все переносится.

$$W(P\nu,P\mu) = W(\Delta P\nu,\Delta P\mu) \leq W(P\Delta\nu,P\Delta\mu)$$

Тогда $W(\nu,\mu) \leq W(\Delta\nu,\Delta\mu) = \sum_{x,y} b_{xy}d(x,y) = \sum_{x,y} b_{xy}W(\delta x,\delta y)$. Если мы применим к $P\Delta\nu$ старый план переносов b_{xy}, а именно будем осуществлять оптимальные переносы из $b_{xy}P\delta x$ в $b_{xy}P\delta y$ по всем x, y, то получим $P\Delta\mu$, однако этот план уже может быть не оптимальным. Получим:

$$W(P\nu,P\mu) \leq W(P\Delta\nu,P\Delta\mu) = W\left(P\left(\sum_x \Delta\nu_x\delta x\right), P\left(\sum_y \Delta\nu_y\delta y\right)\right) =$$

$$= W\left(P\left(\sum_x\left(\sum_y b_{xy}\right)\delta x\right), P\left(\sum_y\left(\sum_x b_{xy}\right)\delta y\right)\right) =$$

$$= W\left(\sum_{x,y} b_{xy}P\delta x, \sum_{x,y} b_{xy}P\delta y\right) \leq$$

$$\leq \sum_{x,y} b_{xy}W(P\delta x,P\delta y) = \sum_{x,y} b_{xy}W(\delta x,\delta y)(1-k_{xy}) =$$

$$= \sum_{x,y} b_{xy}d(x,y)(1-k_{xy}) = \left(\sum_{x,y}\frac{b_{xy}d(x,y)(1-k_{xy})}{W(\nu,\mu)}\right)W(\nu,\mu) =$$

$$= [W(\nu,\mu) = \sum_{x,y} b_{xy} d(x,y)] =$$

$$= \left(1 - \sum_{x,y}\left(\frac{b_{xy}d(x,y)k_{xy}}{\sum b_{xy}d(x,y)}\right)\right) W(\nu,\mu),$$

где k_{xy} - кривизна Риччи между точками x, y.

Коэффициент $\beta_{xy} = \frac{b_{xy}d(x,y)}{\sum b_{xy}d(x,y)}$ имеет смысл доли вклада стоимости переноса массы b_{xy} из x в y в общей оптимальной стоимости переноса $W(\nu,\mu) = \sum b_{xy}d(x,y)$ (при этом $d(x,y)$ играет роль тарифа за перевозку единичной массы из x в y).

Замечание.

В практических задачах, для заданных ν, μ, коэффициенты b_{xy} могут быть найдены численно с использованием стандартных алгоритмов для транспортной задачи [4]. Естественной трудностью при этом становится большой объем вычислений, связанный с комбинаторным характером решаемой задачи.

Следствие 1.

Аналогично получаем:

$$W(P^n\nu, P^n\mu) \leq \sum_{x,y} b_{xy} W(P^n\delta x, P^n\delta y)$$

Следствие 2.

Применив n раз Лемму 1 , получим:

$$W(P^n\nu, P^n\mu) \leq \prod_{i=1}^{n}\left(1 - \sum_{x,y}\left(\frac{b_{xy,i}d(x,y)k_{xy}}{\sum b_{xy,i}d(x,y)}\right)\right) W(\nu,\mu),$$

где $b_{xy,i}$ - оптимальный план переноса на шаге i, обеспечивающий $W(P^{i-1}\nu, P^{i-1}\mu)$.

Замечание.

В свете данной леммы и ее следствий, вычислив численно b_{xy}, мы получаем инструмент для определения скорости сходимости марковского процесса для конкретной прикладной задачи.

Рассмотрим более подробно оптимальный план переброски масс, обеспечивающий $W(\nu, \mu)$. Он не единственный в общем случае. Зафиксируем некоторый такой оптимальный план b_{xy}, означающий, что в оптимальном случае, при трансформировании ν в μ, мы переносим массу b_{xy} из вершины x в вершину y. Стоимостной вклад такого отдельного переноса в $W(\nu, \mu)$ составит $b_{xy}d(x, y)$, а их сумма по всем x, y составит полностью $W(\nu, \mu)$. Однако для дальнейших рассуждений было бы удобно перейти к ситуации, когда переносы масс осуществляются только между соседними точками x', y', $d(x', y') = 1$. То есть, заменить оптимальный план переноса масс b_{xy} эквивалентным ему другим оптимальным планом $a_{x'y'}$, где между соседними точками x', y' переносится масса $a_{x'y'}$, который также обеспечивает $W(\nu, \mu)$ в том смысле, что сумма слагаемых $a_{x'y'}d(x', y') = a_{x'y'}$ по всем $x', y', d(x', y') = 1$ также равна $W(\nu, \mu)$. Чтобы обеспечить возможность этого, следует считать корректной следующую процедуру:

Пусть, например, точки x, y соединяет кратчайший маршрут $x \to x' \to y' \to y, d(x, y) = 3$, тогда из точки x переносим массу b_{xy} на единичное расстояние, затем из точки x' переносим массу b_{xy} в соседнюю точку y' на единичное расстояние, затем из точки y' переносим массу b_{xy} в соседнюю точку y. Таким образом, мы заменили один перенос массы b_{xy} из вершины x в вершину y тремя переносами между соседними точками масс $a_{xx'}, a_{x'y'}, a_{y'y}$, каждая из которых равна b_{xy}. Можно считать эти переносы совершенными одновременно. Если алгоритмизировать этот процесс (с целью проведения численных расчетов для решения прикладных задач), все корректно, поскольку будут осуществляться пошаговые операции. Если в данном примере перенос массы b_{xy} считать единственным, то $W(\nu,\mu) = b_{xy}d(x,y) = a_{xx'} + a_{x'y'} + a_{y'y}$. Поскольку в вершинах x', y' в итоге реализации нового общего плана не прибавилось и не убавилось массы, то перенос масс $a_{xx'}, a_{x'y'}$ мы будем называть транзитным (в том смысле, что масса не остается в вершине окончательно, а переносится далее к следующим вершинам).

В общем случае мы осуществляем декомпозицию - заменяем перенос массы b_{xy} из вершины x в вершину y на расстояние $d(x,y)$ последовательными одношаговыми переносами этой же массы в количестве $d(x,y)$, которые маркируем символами $a_{x'y'}$. Следует отметить, что в общем случае может встретиться ситуация, когда несколько различных элементов $b_{x_1,y_1}, \ldots, b_{x_i,y_i}, \ldots, b_{x_n,y_n}$ исходного оптимального плана таковы, что при декомпозиции фрагмент $\ldots \to x' \to y' \to \ldots$ входит во все выбранные кратчайшие маршруты из x_i в y_i. Тогда $a_{x'y'} = b_{x_1,y_1} + \ldots + b_{x_i,y_i} + \ldots + b_{x_n,y_n}$.

То есть мы должны суммировать все частные переносы, как окончательные (когда перенесенная масса остается в вершине), так и транзитные (когда эта масса переносится далее), осуществляемые на участке $\ldots \to x' \to y' \to \ldots$, возникающие при одношаговой декомпозиции исходного плана.

Визуально это можно ассоциировать со следующим примером:

Пусть автомобили (принадлежащие некоторому грузоперевозчику) в городе перевозят грузы b_{xy} из пунктов x в пункты y оптимальным для грузоперевозчика образом по тарифу, равному километражу маршрута за перевозку единицы груза. Если теперь на каждом километровом участке

$x' \to y'$ на всех улицах города фиксировать все перевозимые грузы и взимать тариф за перевозку, то грузооборот на этом участке составит $a_{x'y'} = b_{x_1,y_1} + \ldots + b_{x_i,y_i} + \ldots + b_{x_n,y_n}$ - сумма грузов всех проехавших через этот участок автомобилей, а полный тарифный сбор на этом участке составит $a_{x'y'}d(x', y') = a_{x'y'}$ и общая оптимальная сумма расходов $W(\nu, \mu)$ на перевозку всех грузов по городу составит с одной стороны сумму $b_{xy}d(x, y)$ по всем x, y, с другой стороны - сумму сборов $a_{x'y'}$ по всем километровым отрезкам графа улиц города.

Корректность описанной процедуры перехода от плана b_{xy} к плану a_{xy} обеспечивается тем, что нам собственно важно выполнение равенства

$$W(\nu, \mu) = \sum_{x,y} b_{xy}d(x, y) = \sum_{x',y':d(x',y')=1} a_{x'y'}d(x', y') = \sum_{x',y':d(x',y')=1} a_{x'y'}$$

Нам достаточно, чтобы фраза "план транспортировки a_{xy}, соответствующий плану b_{xy}"подразумевала вышеописанную процедуру и равенство. Можно также считать, что план транспортировки a_{xy} - это корректный план переносов масс, с той особенностью, что можно в одну и ту же вершину одновременно перенести массу и унести из нее массу, в результате чего в вершине останется разность между принесенной и унесенной массами.

Важно, что оба плана a_{xy}, b_{xy} являются оптимальными, обеспечивающими $W(\nu, \mu)$ и в этом смысле эквивалентны. Можно сказать также, что план a_{xy} ассоциирован с планом b_{xy} или что план a_{xy} является декомпозицией (одношаговой) плана b_{xy}.

Выигрыш в данном случае заключается в том, что, при оценке скорости сходимости марковских процессов, вместо вычисления кривизн между всеми вершинами x, y будет достаточно знать кривизны только между соседними вершинами.

Лемма 2.

Пусть a_{xy} - некоторый план переноса, соответствующий оптимальному плану b_{xy} (из ν в μ), где каждый перенос b_{xy} на расстояние $d(x, y)$ заменен последовательными переносами (в количестве $d(x, y)$) этой массы на единичные расстояния последовательно по одному из кратчайших маршрутов из x в y. То есть, элемент a_{xy} - суммарный объем переноса между соседними вершинами x, y с учетом всех транзитных переносов.

Тогда план переноса a_{xy} из ν в μ также оптимальный.

Доказательство:

Зафиксируем некоторый кратчайший маршрут графа марковской цепи τ_{xy} из вершины x в y. $\tau_{xy} : x = z_{xy}(0) \to z_{xy}(1) \to \ldots \to z_{xy}(d(x, y)) = y$. По свойствам меры, стоимость переноса единичной массы из x в y можно мажорировать суммой стоимостей последовательных транзитных переносов этой массы на единичные расстояния в количестве $d(x, y)$ по цепочке вершин на маршруте, соединяющем x и y в графе марковской цепи.

$$d(x, y) = W(\delta x, \delta y) \leq \sum_{j=1}^{d(x,y)} W(\delta z_{xy}(j-1), \delta z_{xy}(j)) = d(x, y)$$

Следовательно, в данном случае имеет место точное равенство, и утверждение леммы становится очевидным.

Отметим, что при этом $\sum_{x,y:d(x,y)=1} a_{xy} = \sum_{x,y} b_{xy} d(x, y) = W(\nu, \mu)$

Лемма 3.

Имеет место неравенство:

$$W(P\nu, P\mu) \leq \left(1 - \sum_{x,y:d(x,y)=1} \alpha_{xy} k_{xy}\right) W(\nu, \mu),$$

где $\alpha_{xy} = \frac{a_{xy}}{\sum_{d(x,y)=1} a_{xy}}$

<u>**Доказательство:**</u>

По свойствам метрики

$$W(P\delta x, P\delta y) \leq \sum_{j=1}^{d(x,y)} W(P\delta z_{xy}(j-1), P\delta z_{xy}(j)) =$$

$$= \sum_{j=1}^{d(x,y)} (1 - k_{z(j-1)z(j)}) W(\delta z_{xy}(j-1), \delta z_{xy}(j))$$

Тогда, учитывая лемму 1,

$$W(P\nu, P\mu) \leq \sum_{x,y} b_{xy} W(P\delta x, P\delta y) \leq$$

$$\leq \sum_{x,y} b_{xy} \left(\sum_{j=1}^{d(x,y)} (1 - k_{z(j-1)z(j)}) W(\delta z_{xy}(j-1), \delta z_{xy}(j) \right) =$$

$$= \sum_{x,y:d(x,y)=1} a_{xy} (1 - k_{xy}) W(\delta x, \delta y) =$$

$$= \sum_{x,y:d(x,y)=1} a_{xy}(1 - k_{xy}) = \left(\sum_{x,y:d(x,y)=1} \frac{a_{xy}(1 - k_{xy})}{W(\nu, \mu)} \right) W(\nu, \mu) =$$

$$= \left[W(\nu, \mu) = \sum_{x,y} b_{xy} d(x, y) = \sum_{x,y:d(x,y)=1} a_{xy} \right] =$$

$$= \left(1 - \sum_{x,y:d(x,y)=1} \left(\frac{a_{xy}}{\sum a_{xy}} \right) k_{xy} \right) W(\nu, \mu)$$

Отметим, что $\alpha_{xy} = \frac{a_{xy}}{\sum_{d(x,y)=1} a_{xy}}$ имеет смысл доли массы (и стоимости, т.к. тариф $d(x,y)=1$), переносимой из вершины x в соседнюю вершину y, в том числе транзитно, при оптимальном по стоимости переносе из ν в μ (где переносы масс разрешены последовательно только на единичные расстояния между соседними вершинами) по отношению к общей стоимости перенесенной массы $W(\nu,\mu)$.

Следствие 1.

Из полученного неравенства, в частности, следует ранее использованный результат, что, если все кривизны между соседними вершинами k_{xy} мажорированы некоторым ε ($k_{x,y} \geq \varepsilon \quad \forall x,y$), то $W(P\nu, P\mu) \leq (1-\varepsilon)W(\nu,\mu)$.

Следствие 2.

Рассмотрим соседей x,y, где $y = x_{rs} = (\ldots, x_r+1, \ldots, x_s-1, \ldots)$. Обозначим $\Omega_{01} = \{(x,y) | y = x_{rs}, x_r = 0, x_s = 1, r \neq s\}$,

$\Omega_{11} = \{(x,y) | y = x_{rs}, x_r = 0, x_s \geq 2 \quad \text{или} \quad x_r \geq 1, x_s = 1, r \neq s\}$.

Тогда, по Лемме 0, имеет место неравенство:

$$W(P\nu, P\mu) \leq [1 - \sum_{x,y:(x,y)\in\Omega_{01}} \left(\frac{a_{xy}}{\sum_{x,y:d(x,y)=1} a_{xy}}\right) \frac{1}{N} -$$

$$- \sum_{x,y:(x,y)\in\Omega_{11}} \left(\frac{a_{xy}}{\sum_{x,y:d(x,y)=1} a_{xy}}\right) \frac{1}{N(N-1)}] W(\nu,\mu)$$

Замечание.

В практических задачах коэффициенты a_{xy} также могут быть найдены численно с использованием алгоритмов, основанных на методах решения транспортной задачи [4]. В этом случае мы также получаем инструмент для определения скорости сходимости марковского процесса.

Некоторый интерес представляет вопрос о том, какую долю занимает количество разных (по составу) пар соседних вершин, обладающих достаточно высокими кривизнами, в общем количестве всех вообще разных пар соседних вершин.

Задача 1.

Пусть Ω - множество всех различных неупорядоченных пар соседних вершин.

Через Ω_1 обозначим множество разных пар вершин, кривизна между которыми становится ненулевой за один шаг марковской цепи. Это пары вершин с кривизнами $\frac{1}{N}$ и $\frac{1}{N(N-1)}$. Через ω_1 обозначим долю вышеуказанных разных пар в общем количестве разных пар соседних вершин, $\omega_1 = \frac{Card\Omega_1}{Card\Omega}$. Через γ_1 обозначим долю пар вершин кривизны $\frac{1}{N}$ в составе Ω_1. Доля пар вершин, обладающих кривизной $\frac{1}{N}$, составит $\gamma_1\omega_1$, а доля пар вершин, обладающих кривизной $\frac{1}{N(N-1)}$ составит $(1-\gamma_1)\omega_1$ в общем количестве разных соседних вершин. Тогда имеет место:

$$\gamma_1 = \frac{N-2}{2M+N-4},$$

$$\omega_1 \geq 1 - \frac{(M-1)(M-2)}{(M+N-4)(M+N-5)},$$

Доказательство:

Обозначим $\omega_i = \dfrac{Card(\Omega_i)}{Card(\Omega)}$ - доли пар вершин с фиксированной кривизной за i шагов в их общем количестве. Сначала найдем число различных векторов $(x_1, \ldots, x_n)$, где $\sum x_i = m$. Значения x_i представимы в виде соответствующего количества "1"и разделители между x_i, x_j - в виде "0". Тогда это соответствует числу векторов $(11\ldots101\ldots10\ldots1)$, содержащих m единиц и $(n-1)$ нулей. Таких векторов всего C^m_{m+n-1} (по аналогии со схемой Бернулли - m успехов (1 - успех) из $(m+n-1)$ испытаний (0 - неуспех)).

Рассмотрим пару вершин $x, y = x_{rs} = (\ldots, x_r+1, \ldots, x_s-1, \ldots)$. Зафиксируем r, s. Запишем виды векторов, их общее количество и соответствующие кривизны для подмножества $\Omega_{1,rs}$.

Первая подгруппа:

$(\ldots,0,\ldots,1,\ldots)$	$C^{M-1}_{(M-1)+(N-3)}$	$k_{01} = \frac{1}{N}$
$(\ldots,0,\ldots,2,\ldots)$	$C^{M-2}_{(M-2)+(N-3)}$	$k_1 = \frac{1}{N(N-1)}$
$(\ldots,0,\ldots,3,\ldots)$	$C^{M-3}_{(M-3)+(N-3)}$	
$\ldots$	$\ldots$	
$(\ldots,0,\ldots,M,\ldots)$	$C^0_{N-3} = 1$	

Вторая подгруппа:

$(\ldots,1,\ldots,1,\ldots)$	$C^{M-2}_{(M-2)+(N-3)}$	$\sim \frac{1}{N(N-1)}$
$(\ldots,2,\ldots,1,\ldots)$	$C^{M-3}_{(M-3)+(N-3)}$	
$(\ldots,3,\ldots,1,\ldots)$	$C^{M-4}_{(M-4)+(N-3)}$	
$\ldots$	$\ldots$	
$(\ldots,M-1,\ldots,1,\ldots)$	$C^0_{N-3} = 1$	

Рассмотрим $C^0_{(N-3)} = 1; C^1_{1+(N-3)} = N-2; C^2_{2+(N-3)} = \frac{(N-1)(N-2)}{2}; \ldots$
Обратим внимание на следующую закономерность:

$$\sum_{i=0}^{m} C^i_{i+(N-3)} = C^{(m+1)}_{(m+1)+(N-3)} \cdot \frac{(m+1)}{(N-2)}$$

Тогда

$$Card(\Omega_{1,rs}) = C^M_{M+(N-3)} \cdot \frac{M}{(N-2)} + C^{M-1}_{(M-1)+(N-3)} \cdot \frac{(M-1)}{(N-2)} =$$

$$= \frac{(M+N-3)(M+N-4)\cdot\ldots\cdot(N-2)}{M!}\cdot\frac{M}{(N-2)}+$$

$$+\frac{(M+N-4)\cdot\ldots\cdot(N-2)}{(M-1)!}\cdot\frac{(M-1)}{(N-2)}=$$

$$=\left(\frac{M+N-3}{(M-1)}+1\right)\cdot\frac{(M+N-4)\cdot\ldots\cdot(N-1)}{(M-2)!}=$$

$$=(2M+N-3)\cdot\frac{(M+N-4)\cdot\ldots\cdot(N-1)}{(M-1)!}$$

Рассмотрим $\Omega_{2,rs}$ за 2 шага:

$(\ldots,1,\ldots,2,\ldots)$	$C^{M-3}_{(M-3)+(N-3)}$	$\sim \frac{1}{N^2(N-1)}$
$(\ldots,1,\ldots,3,\ldots)$	$C^{M-4}_{(M-4)+(N-3)}$	
$\ldots$	$\ldots$	
$(\ldots,1,\ldots,M-1,\ldots)$	$C^{0}_{N-3}=1$	

$(\ldots,2,\ldots,2,\ldots)$	$C^{M-4}_{(M-4)+(N-3)}$	$\sim \frac{1}{N^2(N-1)}$
$(\ldots,3,\ldots,2,\ldots)$	$C^{M-5}_{(M-5)+(N-3)}$	
$\ldots$	$\ldots$	
$(\ldots,M-2,\ldots,2,\ldots)$	$C^{0}_{N-3}=1$	

$$Card(\Omega_{2,rs})=C^{M-2}_{(M-2)+(N-3)}\cdot\frac{(M-2)}{(N-2)}+C^{M-3}_{(M-3)+(N-3)}\cdot\frac{(M-3)}{(N-2)}=$$

$$=\left(\frac{M+N-5}{M-3}+1\right)\cdot\frac{(M+N-6)\cdot\ldots\cdot(N-1)}{(M-4)!}=$$

$$=(2M+N-8)\cdot\frac{(M+N-6)\cdot\ldots\cdot(N-1)}{(M-3)!}$$

$$\ldots$$

$$Card(\Omega_{l,rs}) = \frac{(M+N-4l)(M+N-2l-2)\cdot\ldots\cdot(N-1)}{(M-2l+1)!}, \qquad \left(l < \frac{M}{2}\right)$$

$$\ldots$$

$$Card(\Omega_{\frac{M}{2},rs}) = N$$

$$\frac{Card(\Omega_{rs})}{Card(\Omega_{1,rs})} = \frac{\sum_{i=1}^{\frac{M}{2}} Card(\Omega_{i,rs})}{Card(\Omega_{1,rs})} = 1 + \frac{(2M+N-8)}{2M+N-4}\cdot\frac{(M-1)(M-2)}{(M+N-4)(M+N-5)} +$$

$$+\frac{(2M+N-12)}{(2M+N-4)}\cdot\frac{(M-1)(M-2)(M-3)(M-4)}{(M+N-4)(M+N-5)(M+N-6)(M+N-7)} + \ldots \leq$$

$$\leq \left[q_1 = \frac{(M-1)(M-2)}{(M+N-4)(M+N-5)}\right] \leq 1 + q_1 + q_1^2 + \ldots = \frac{(1-q_1^{\frac{M}{2}})}{1-q_1} \leq \frac{1}{1-q_1}$$

Тогда

$$\omega_1 = \frac{Card(\Omega_1)}{Card(\Omega)} = \frac{Card(\Omega_{1,rs})}{Card(\Omega_{rs})} \geq 1 - q_1 = 1 - \frac{(M-1)(M-2)}{(M+N-4)(M+N-5)}$$

Выделим из состава $\Omega_{1,rs}$ долю пар вершин кривизны $\frac{1}{N}$

$$\gamma_1 = \frac{C^{M-1}_{(M-1)+(N-3)}}{\Omega_{1,rs}} = \frac{\left(\frac{(M+N-4)\cdot\ldots\cdot(N-2)}{(M-1)!}\right)}{\left(\frac{(2M+N-4)(M+N-4)\cdot\ldots\cdot(N-1)}{(M-1)!}\right)}$$

$$\gamma_1 = \frac{N-2}{2M+N-4}; \qquad (1-\gamma_1) = \frac{2M-2}{2M+N-4}$$

Доказано.

Следствие 1.

Для больших $M, N \gg 1$ доля пар вершин, обладающих кривизной $\frac{1}{N}$, составит в общем количестве разных пар соседних вершин величину

$\gamma_1\omega_1 \approx \left(\frac{N}{N+M}\right)^2$, а доля пар вершин, обладающих кривизной $\frac{1}{N(N-1)}$ составит $(1-\gamma_1)\omega_1 \approx \frac{2NM}{(N+M)^2}$, совместная доля вершин кривизн $\frac{1}{N}$, $\frac{1}{N(N-1)}$ составит $\omega_1 \approx 1 - \frac{1}{(1+\frac{N}{M})^2}$

Доказательство:

$$\text{При} \quad M, N \gg 1 \quad \omega_1 \approx 1 - \frac{M^2}{(M+N)^2} = 1 - \left(\frac{1}{1+\frac{N}{M}}\right)^2$$

$$\gamma_1 \approx \tfrac{N}{2M+N}, \quad (1-\gamma_1) \approx \tfrac{2M}{2M+N}$$

$$\gamma_1\omega_1 \approx \frac{N}{2M+N}\left(1 - \left(\frac{M}{M+N}\right)^2\right) = \frac{N(2MN+N^2)}{(2M+N)(M+N)^2} = \left(\frac{N}{M+N}\right)^2$$

$$(1-\gamma_1)\omega_1 \approx \frac{2M}{2M+N} \cdot \left(1 - \left(\frac{M}{M+N}\right)^2\right) = \frac{2M(2MN+N^2)}{(2M+N)(M+N)^2} = \frac{2MN}{(M+N)^2}$$

В качестве гипотезы выдвигаем предположение о том, что скорость сходимости в исследуемой (N, M) задаче коррелированна с относительной долей пар вершин высокой кривизны (и вообще с набором относительных долей пар вершин всех кривизн).

Для сопоставления расчетов скорости сходимости в практических задачах, имеющих различные начальные данные, предлагаем сконструировать некоторый маркер, который можно назвать фиктивной кривизной k^*, являющийся параметром (N, M) задачи.

Итак, предлагаем ввести понятие фиктивной кривизны k^*, являющейся некоторым характеристическим параметром (N, M) задачи, следующим формальным образом:

Для k, определяемого равенством $W(P\nu, P\mu) = (1-k)W(\nu, \mu)$, по Лемме 3 имеем

$$k = \left(1 - \sum_{x,y:d(x,y)=1} \left(\frac{a_{xy}}{\sum a_{xy}}\right) k_{xy}\right) W(\nu, \mu)$$

Если теперь формально положить $a_{xy} = \frac{1}{Card(\Omega)} \quad \forall x, y, d(x,y) = 1$, то по вышеприведенной формуле мы и получим k^*. Такая инициация является формальной, так как на практике она означала бы, что через все ребра, соединяющие все соседние вершины, при оптимальном переносе масс переносятся равные массы $\frac{1}{Card(\Omega)}$. Но на практике это нереализуемо, такого примера не существует (это связано с тем, что некоторые соседние вершины образуют "треугольники"и попытка переноса равных масс по всем трем ребрам, не важно в каком направлении, ведет к потере оптимальности, одно ребро в "треугольнике"обязательно должно быть свободно от переноса массы). То есть фиктивная кривизна не является кривизной в действительном понимании, так как нет примера, где она корректно вычисляется. Тем не менее, этот параметр может быть полезен, как некоторая характеристика (N, M) задачи.

Лемма 4.

Для параметра фиктивной кривизны k^* имеет место следующая оценка:

$$k^* \geq \gamma_1 \omega_1 \frac{1}{N} + (1 - \gamma_1) \omega_1 \frac{1}{N(N-1)},$$

где γ_1, ω_1 взяты из формулировки Задачи 1.

Доказательство:

Следует непосредственно из Следствия 2 Леммы 3 и Задачи 1.

Следствие 1.

Для больших $M, N \gg 1$ имеет место

$$k^* \approx \frac{(N^2+2M)}{(M+N)^2} \cdot \frac{1}{N}$$

<u>Доказательство:</u>

По Следствию 1 Задачи 1 и Лемме 4 получим:

$$k^* \approx \left(\frac{N}{N+M}\right)^2 \cdot \frac{1}{N} + \frac{2MN}{(M+N)^2} \cdot \frac{1}{N(N-1)} \approx$$
$$\approx \frac{(N^2+2M)}{(M+N)^2} \cdot \frac{1}{N}$$

Так как на каждом шаге i марковской цепи сжимающий множитель $(1-k_i)$ может меняться, то предлагаем ввести понятие средней кривизны марковской цепи за n шагов следующим образом:

$$\bar{k}_n = 1 - \sqrt[n]{\prod_{i=1}^{n}(1-k_i)} = 1 - \sqrt[n]{\frac{W(P^n\nu, P^n\mu)}{W(\nu,\mu)}}$$

Проиллюстрируем рассмотренные понятия и процессы на следующем элементарном примере:

<u>Пример 1.</u>

Рассмотрим задачу $(N=2, M=1)$ - два игрока, три монеты. Вершины $1,2,3,4$ - это $(3,0),(2,1),(1,2),(0,3)$ соответственно. Соседними парами вершин являются $\{1,2\},\{2,3\},\{3,4\}$.

Матрица переходов P $= \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix}$

Стационарное распределение $\pi = (0.25, 0.25, 0.25, 0.25)$. Строка i матрицы переходов соответствует $P\delta_i$. Кривизны между соседними точками равны:

$k_{12} = 0.5, k_{23} = 0, k_{34} = 0.5$

Тогда фиктивная кривизна $k^* = \frac{2}{3} \cdot 0.5 + \frac{1}{3} \cdot 0$

Для заданного ν возьмем $\mu = P\nu$, тем самым мы будем сравнивать последующий шаг с предыдущим: $W(\nu, \mu) = W(\nu, P\nu)$.

В данном примере достаточно несложно рассчитывать $W(\nu, P\nu)$, результаты некоторых расчетов представляем в следующих таблицах:

Номер шага цепи n	$P^n\nu$	$W(P^{n-1}\nu, P^n\nu)$	$\bar{k}_{n-1}$
0	$0, 0.3, 0, 0.7$		
1	$0.15, 0, 0.5, 0.35$	0.65	
2	$0.075, 0.325, 0.175, 0.425$	0.4	0.3846
3	$0.2, 0.125, 0.375, 0.3$	0.325	0.2929
4	$0.1625, 0.2875, 0.2125, 0.3375$	0.2	0.3248
5	$0.225, 0.1875, 0.3125, 0.275$	0.1625	0.2929
...	...	...	...
97	$0.25, 0.25, 0.25, 0.25$	$1.44E - 15$	0.2938
98	$0.25, 0.25, 0.25, 0.25$	$1.17E - 15$	0.2929

Номер шага цепи n	$P^n\nu$	$W(P^{n-1}\nu, P^n\nu)$	$\bar{k}_{n-1}$
0	$0.4, 0, 0, 0.6$		
1	$0.2, 0.2, 0.3, 0.3$	0.5	
2	$0.2, 0.25, 0.25, 0.3$	0.05	0.9
3	$0.225, 0.225, 0.275, 0.275$	0.05	0.6838
4	$0.225, 0.25, 0.25, 0.275$	0.25	0.6316
5	$0.2375, 0.2375, 0.2625, 0.2625$	0.25	0.5271
...	...	...	...
97	$0.25, 0.25, 0.25, 0.25$	$7.22E-16$	0.30749
98	$0.25, 0.25, 0.25, 0.25$	$3.61E-16$	0.30989

ν	$\bar{k}_1$	$\bar{k}_5$	$\bar{k}_{90}$
$1, 0, 0, 0$	0.5	0.3402	0.2929
$0, 1, 0, 0$	0.25	0.2845	0.2929
$0, 0, 1, 0$	0.25	0.2845	0.2929
$0, 0, 0, 1$	0.5	0.3402	0.2929

Для данного примера наблюдается наличие предельных значений средних кривизн, которые приближенно согласуются с фиктивной кривизной.

Список литературы

[1] M. Erbar, J. Maas. *Ricci Curvature of Finite Markov Chains via Convexity of the Entropy.* Archive for Rational Mechanics and Analysis vol. 206 issue 3 December 2012. p. 997 - 1038

[2] Y. Ollivier. *Ricci curvature of Markov chains on metric spaces.* J. Funct. Anal., 256(3):810Ц864, 2009.

[3] Villani, C.: *Topics in Optimal Transportation.* Am. Math. Soc., Providence (2003)

[4] Кремер Н.Ш. *Исследование операций в экономике.* - М.: ЮНИТИ, 2006.