Записки научных
семинаров ПОМИ
Том 411, 2013 г.

А. М. Вершик

ДВА СПОСОБА ОПРЕДЕЛЕНИЯ СОГЛАСОВАННЫХ МЕТРИК НА СИМПЛЕКСЕ МЕР

**К столетию Л. В. Канторовича
и к 70-летию выхода его статьи о метрике**

§1. Согласованные метрики

Пусть (X, r) – сепарабельное метрическое пространство; обозначим через $\mathrm{Meas}\,(X)$ симплекс вероятностных борелевских мер на (X, r) и снабдим его слабой топологией двойственности с пространством непрерывных функций с компактным носителем. Меру, сосредоточенную в точке, обозначим, как обычно, δ_x, и положим $\varepsilon_{x,y} = \delta_x - \delta_y$. Мы будем рассматривать различные метрики на $\mathrm{Meas}\,(X)$.

Определение 1. *Будем называть метрику ρ на симплексе вероятностных мер согласованной с метрикой r на пространстве X (или просто согласованной, если ясно, о какой метрике r идет речь), если*

$$\rho_r(\delta_x, \delta_x) = r(x, y).$$

Все согласованные метрики образуют выпуклое множество в пространстве всех метрик как вещественных функций на пространстве $X \times X$. Во многих ситуациях, встречающихся в функциональном анализе, оптимизации и математической экономике, теории вероятностей, статистике, теории динамических систем и т.д., возникает необходимость в рассмотрении той или иной согласованной метрики, т.е. в некотором "поднятии" метрики с пространства на меры на этом пространстве. Диапазон согласованных метрик огромен, и выбор той или иной объясняется особенностями конкретной задачи. Однако имеет смысл исследовать некоторые общие свойства таких метрик и их совокупности в целом. Мы приводим два различных – двойственных – способа их определения, которые хотя и не исчерпывают все возможности, но включают наиболее существенные из них.

Ключевые слова: метрика Канторовича, согласованные метрики, трансляционная инвариантность, транспортные задачи.

Работа поддержана грантом РФФИ 11-01-00677-а.

Заметим, что любую метрику на симплексе вероятностных мер можно продолжить по однородности на симплексы мер с заданным значением меры на всем пространстве: $\rho_r(\lambda\mu_1, \lambda\mu_2) = \lambda\rho_r(\mu_1, \mu_2)$, $\lambda \in \mathbb{R}$. Если принять условие $\rho(\lambda_1\mu_1, \lambda_2\mu_2) = |\lambda_1 - \lambda_2| + \rho(\mu_1, \mu_2)$, то мы получим метрику на конусе всех неотрицательных мер. Но удобнее рассматривать пространство знакопеременных мер с нулевой общей массой, тогда первое слагаемое исчезает.

Выделим особо одно важное свойство метрик на симплексе.

Определение 2. *Метрика ρ_r на симплексе вероятностных мер называется трансляционно инвариантной, если выполнено следующее равенство:*

$$\rho_r(\lambda\mu_1 + (1-\lambda)\mu_0, \lambda\mu_2 + (1-\lambda)\mu_0) = \lambda\rho_r(\mu_1, \mu_2),$$

где $\mu_0, \mu_1, \mu_2 \in \mathrm{Meas}\,(X)$, $0 \leqslant \lambda \leqslant 1$.

Трансляционно инвариантная метрика на симплексе продолжается до трансляционно инвариантной метрики на всем конусе неотрицательных мер:

$$\rho(\theta_1 + \theta_0, \theta_2 + \theta_0) = \rho(\theta_1, \theta_2);$$

здесь θ_0, θ_1, θ_2 – произвольные неотрицательные меры.

§2. ПРЯМОЙ МЕТОД: МИНИМУМ ТРАНСПОРТНЫХ ЗАТРАТ

Пусть заданы две вероятностные меры μ_1 и μ_2 на метрическом пространстве (X, r). Определим выпуклое множество Ψ_{μ_1,μ_2} борелевских мер ψ на $X \times X$, проекции (маргинальные меры) которых на первую и вторую компоненты суть μ_1 и μ_2. Это так называемые бистохастические меры (мы употребляем этот термин в более широком, чем обычно, смысле; обычно его используют, если обе проекции равны). Каждая бистохастическая мера определяет *полиморфизм пространства с мерой* (X, μ_1) *на пространство с мерой* (X, μ_2) (метрика в этом определении не участвует). Другие названия для того же понятия – coupling, joining, мера Юнга, марковское отображение, бирасслоение, соответствие и т.д. (см. [13]).

С вероятностной точки зрения, бистохастическая мера есть двумерное совместное распределение двух случайных величин с распределениями μ_1 и μ_2. В эконометрической интерпретации меры ψ суть *планы перевозок*, а множество Ψ_{μ_1,μ_2} есть множество всех планов, реализующих спрос.

Выберем какое-либо нормированное или метрическое пространство измеримых функций (L^p, $p \geqslant 1$, $p = \infty$, пространство Орлича и др.) на пространстве с мерой (Z, ν); обозначим соответствующую норму или метрику (в этом случае имеется в виду расстояние до нуля) через $\mathcal{N}_{(Z,\nu)}$ или $\mathcal{N}_\nu$, если пространство Z подразумевается.

Определение 3. *Определим согласованную метрику, соответствующую норме (метрике) $\mathcal{N}$ на множестве мер* $\operatorname{Meas}(X)$, *следующим образом:*

$$\rho^r_{\mathcal{N}}(\mu_1, \mu_2) = \inf_{\psi \in \Psi_{\mu_1,\mu_2}} \mathcal{N}_\psi(r(\cdot,\cdot))$$

в тех случаях, когда метрика $r(\cdot,\cdot)$ *принадлежит соответствующему $\mathcal{N}$ нормированному (метрическому) пространству измеримых функций; в остальных случаях метрика $\rho_{\mathcal{N}}$ не определена.*

Основной пример (L^p-аналог метрики Канторовича) таков. Предположим, что метрика r интегрируема по мерам μ_1, μ_2:

$$\int\limits_{X \times X} r(x,y) d\mu_1(x)\mu_2(y) < \infty.$$

Выберем в качестве нормы L^p-норму с $p \geqslant 1$; тогда данное выше определение приводит к следующему классу метрик:

$$\rho^p_r(\mu_1, \mu_2) = \inf_{\psi \in \Psi_{\mu_1,\mu_2}} \left\{ \int\limits_{X \times X} r(x,y)^p \, d\psi(x,y) \right\}^{1/p}.$$

При $p = 1$ мы получаем классическую метрику Канторовича $k_r \equiv \rho^1_r$:

$$k_r(\mu_1, \mu_2) = \inf_{\psi \in \Psi_{\mu_1,\mu_2}} \int\limits_{X \times X} r(x,y) \, d\psi(x,y).$$

Особое положение занимает здесь случай $p = \infty$:

$$\rho^\infty_r(\mu_1, \mu_2) = \inf_{\psi \in \Psi_{\mu_1,\mu_2}} \operatorname{esssup}_{(x,y) \in (X \times X, \psi)} r(x,y).$$

Эта метрика есть не что иное, как широко известная метрика Леви–Прохорова (см. ниже).

Оптимизационный смысл этих поднятий метрики с метрического пространства на симплекс мер ясен: расстоянием между мерами объявляется минимум (или, точнее, инфимум) норм метрики как измеримой функции на квадрате пространства с двумерным распределением, имеющим заданные маргинальные распределения.

Исключительная роль нормы L^1, т.е. метрики Канторовича, состоит в следующем ее свойстве.

Предложение 1. *Единственная норма $\mathcal{N}$, для которой метрика $\rho^r_{\mathcal{N}}$ является трансляционно инвариантной, есть норма L^1. Следовательно, в предложенном классе метрик только метрика Канторовича трансляционно инвариантна. Метрика Канторовича не превышает все остальные нормы этого класса: $k_r(\mu_1, \mu_2) \leqslant \rho^r_{\mathcal{N}}(\mu_1, \mu_2)$ для всех мер μ_1, μ_2, для которых расстояния определены.*

Эконометрический смысл этого свойства прозрачен: если добавить к начальному и конечному распределению масс одно и то же распределение, то затраты от этого не изменятся и перевозки можно будет осуществлять с помощью того же оптимального плана, что и без добавления меры. Если же этого свойства нет, то окажется выгодным использовать в качестве промежуточной меры добавленную меру. Это обстоятельство особенно ясно для L^p-норм при $p > 1$: положительная степень метрики, вообще говоря, метрикой не является, и неравенство треугольника меняется на противоположное. Метрика Канторовича соответствует постановке *линейных транспортных задач.* Другие же метрики этого класса связаны с нелинейными задачами.

Метрики типа L^p, или степенные метрики Канторовича, интенсивно использовались в последние годы в теории оптимальных перевозок; см. [15] и цитированную там литературу. Однако трудно сейчас сказать, кто ввел их первым, и в любом случае неправильно называть их метриками Вассерштейна, который их не рассматривал (подробнее об этом см. в [14]).

Среди только что введенных метрик заслуживает внимания еще одна, а именно соответствующая L^∞.

Предложение 2. *L^∞-аналог метрики Канторовича совпадает с метрикой Леви–Прохорова.*

Удивительно, что несложное доказательство этого факта по существу было дано В. Штрассеном [12] еще в 1965 году, но, видимо, аудитория, интересующаяся этой метрикой, не знала ничего об обобщениях метрики Канторовича, и наоборот. А именно, Штрассен (теорема 11 в [12]) заметил, что метрика Леви–Прохорова может быть определена ровно по той формуле, которая приведена выше; ее словесная формулировка такова: расстояние Леви–Прохорова равно минимальной ширине полоски около диагонали, в которую можно поместить бистохастическую меру с данными маргинальными распределениями.

Подробнее, введем число, характеризующее бистохастичесую меру:

$$\mathcal{N}_\psi(r(\cdot,\cdot)) \equiv \min\{\epsilon : \psi\{(x,y) : r(x,y) < \epsilon\} = 1\},$$

если такое конечное ϵ существует. Тогда расстояние Леви–Прохорова имеет такой же вид, как и метрики выше:

$$\rho_{LP}(\mu_1,\mu_2) = \inf_{\psi\in\Psi_{\mu_1,\mu_2}} \mathcal{N}_\psi(r(\cdot,\cdot)).$$

Разумеется, это расстояние мажорирует все p-метрики Канторовича при $p < \infty$; оно конечно далеко не для всех пар мер.

Следует особо выделить случай $p = 2$, т.е. L^2-аналог метрики Канторовича; его систематически рассматривал Ф. Отто ("исчисление Отто"; см., например, [10]). Основное предположение состоит в том, что симплекс мер на римановом многообразии, снабженный L^2-аналогом метрики Канторовича, является бесконечномерным римановым многообразием с рядом замечательных свойств, отчасти наследующих свойства самого риманова многообразия. Эта метрика связана с уравнением Монжа–Ампера.

По такому же образцу можно строить совместимые метрики, порожденные не нормами, а другими функционалами в пространствах измеримых функций, включая ненормируемые и даже не локально выпуклые пространства. Общее свойство всех совместимых метрик этого класса состоит в том, что расстояние между мерами по каждой из них есть инфимум по всем бистохастическим мерам с заданными маргинальными мерами некоторого функционала от метрики на исходном пространстве и этот функционал определен единым образом (функториально для всех бистохастических мер) на некотором пространстве измеримых функций. Никаких общих симметрий все совместимые метрики не имеют. В следующем разделе мы определим другой класс совместимых метрик; они, наоборот, объединены некоторым общим условием симметрии.

§3. Двойственный метод (максимум потенциала) и трансляционно инвариантные метрики

Рассмотрим другой способ определения согласованной метрики. Для этого напомним основную теорему работы [5] о двойственном определении метрики Канторовича.

Теорема 1. *Метрика k_r может быть определена следующим образом:*

$$k_r(\mu_1, \mu_2) = \sup_{u \in Lip_1(X,r)} \int\limits_X u(x)\, d(\mu_1 - \mu_2)(x),$$

где супремум берется по всем функциям, удовлетворяющим условию Липшица с константой единица: $|u(x) - u(y)| \leqslant r(x, y)$. Супремум достигается на некоторой липшицевой функции u_0, которая называется потенциалом в данной задаче; при этом в первом определении метрики k_r (см. выше) инфимум также достигается на некоторой мере (плане перевозок) ψ_0, которая называется потенциальным планом; связь этих двух задач состоит в том, что

$$\psi_0\{(x, y) : u_0(x) - u_0(y) = r(x, y)\} = 1, \quad u_0(x) - u_0(y) \leqslant r(x, y);$$

последнее равенство есть условие оптимальности плана ψ_0 и потенциала u_0. Совпадение двух определений метрики Канторовича есть то, что называется теоремой двойственности для данной пары двойственных задач.[1]

Появление липшицевых функций здесь совершенно естественно: определяющие их неравенства возникают как неравенства в двойственной задаче. Но наиболее четкая трактовка их появления связана с нормой Канторовича–Рубинштейна, см. далее.

Заметим, что теорема, сформулированная выше, доказана в [5] для компактного метрического пространства (X, r); она остается верной для произвольного сепарабельного метрического пространства, но меры μ_1, μ_2 должны быть подчинены условию интегрируемости метрики, выписанному выше.

Для дальнейшего удобно ввести новые банаховы пространства, ассоциированные с трансляционно инвариантными метриками. Пусть (X, r) – сепарабельное метрическое пространство и ρ_r – какая-либо

[1] Это был первый пример теоремы двойственности для бесконечномерных задач линейного программирования; общая теорема, в тех случаях, когда она имеет место, формулируется по тому же образцу

трансляционно инвариантная согласованная с r метрика на симплексе мер. Как было сказано, она продолжается до трансляционно инвариантной метрики на конусе неотрицательных мер. Но поскольку произвольная мера конечной вариации единственным образом представляется в виде разности двух неотрицательных мер, можно определить норму в пространстве знакопеременных мер конечной вариации, полагая

$$\|\mu_1 - \mu_2\| = \rho_r(\mu_1, \mu_2);$$

здесь ρ_r считается уже продолженной на конус по формуле, приведенной ранее. Это определение, в силу однородности и трансляционной инвариантности, корректно задает норму в пространстве мер. Норма определена во всяком случае для дискретных мер с конечным носителем. Таким образом, построено банахово пространство, связанное с метрикой ρ_r. Такие нормы мы называем нормами, *согласованными с метрикой*. Несколько в иной форме это построение проведено в [9].

Исходя из метрики Канторовича, мы приходим к норме Канторовича–Рубинштейна $\|\cdot\|_{\mathrm{KR}}$. Она определена в [7] для пространства всех борелевских мер конечной вариации на компакте. Определение корректно и для произвольного метрического пространства, но уже не для всех мер. Однако норма определена во всяком случае для дискретных мер с конечным носителем, а затем ее область определения можно получить с помощью пополнения по норме. Нужно иметь в виду, что уже даже для компакта (и даже для отрезка $[0,1]$) это пополнение содержит не только меры, но и более сложные объекты (типа распределений Шварца), однако это для нас несущественно. Замечательно, что сопряженным пространством (пространством всех линейных непрерывных функционалов) к нормированному пространству Канторовича–Рубинштейна является пространство липшицевых функций. Это обстоятельство было впервые отмечено в работе [7], и именно оно объясняет появление этих функций в двойственной задаче (см. выше).

Любая согласованная с метрикой r трансляционно инвариантная метрика ρ_r на симлексе задает согласованную полунорму на пространстве $V_0(X)$ дискретных знакопеременных мер с конечным носителем, т.е., попросту говоря, на пространстве конечных формальных вещественных линейных комбинаций точек метрического пространства, по формуле

$$\|\mu_1 - \mu_2\| = \rho_r(\mu_1, \mu_2),$$

где

$$\mu_1 = \sum_{i=1}^{n} c_i \delta_{x_i}, \quad \mu_2 = \sum_{j=1}^{m} d_j \delta_{y_j}, \quad c_i, d_j > 0, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m.$$

Факторизация пространства $V_0(X)$ по нулевому подпространству этой полунормы порождает нормированное пространство, в которое изометрично вкладывается наше метрическое пространство. И наоборот, всякое изометричное вложение данного метрического пространства в банахово пространство определяется с помощью согласованной (полу)нормы на пространстве формальных линейных комбинаций точек пространства. Все это подробно обсуждается в работе [9]. Для нас здесь важен следующий простой факт, который, видимо, впервые отмечен в указанной работе.

Теорема 2 ([9]). *Норма Канторовича–Рубинштейна максимальна среди всех согласованных норм, и, в частности, метрика Канторовича является максимальной среди всех трансляционно инвариантых метрик. Пополнение пространства дискретных мер с конечным носителем по норме Канторовича–Рубинштейна является наименьшим банаховым пространством из всех полученных пополнением по согласованным нормам.*

Доказательство вытекает из приведенной теоремы двойственности, точнее из двойственного определения нормы Канторовича–Рубинштейна. Иначе говоря, это определение налагает минимальные ограничения на двойственные переменные – лишь условия Липшица. Всякая другая норма содержит дополнительные ограничения на двойственные переменные. Можно сказать, что сопряженные пространства ко всем остальным нормированным пространствам в этой конструкции содержатся в пространстве липшицевых функций. Уже в конечномерном случае любопытна геометрия единичного шара в норме Канторовича–Рубинштейна: он является выпуклой оболочкой векторов $\frac{\epsilon_{x,y}}{r(x,y)}$ и, тем самым, наименьшим среди шаров, порожденных совместимыми трансляционно инвариантными метриками. Для однородной метрики на конечном множестве из n точек этот шар есть выпуклая оболочка корней в подалгебре Картана алгебры Ли серии A_n.

Следствие 1. *Всякая трансляционно инвариантная согласованная метрика может быть определена следующим образом:*

$$\rho^r_{\mathcal{L}}(\mu_1,\mu_2) = \sup_{u\in\mathcal{L}} \left| \int u(x) d(\mu_1-\mu_2) \right| \equiv \|\mu_1-\mu_2\|_{\mathcal{L}},$$

где $\mathcal{L}$ есть подходящее линейное подпространство липшицевых функций на метрическом пространстве (X,r).

Это есть двойственный способ построения согласованных норм. Заметим, что, по-видимому, из всех согласованных метрик только метрика Канторовича может быть определена обоими способами. Как уже упоминалось, все остальные метрики, полученные первым способом, вообще говоря, не являются трансляционно инвариантными и не могут определять норм.

Замечательно, что и по существу своих применений эти классы сильно отличаются. Во второй (трансляционно инвариантный) класс входят, кроме метрики Канторовича, метрика Хаусдорфа–Куратовского, так называемая двухточечная метрика, играющая очень важную роль в теории линейно жестких пространств (см. [9]), и др.

Мы заключим статью следующим парадоксальным утверждением: существуют такие метрические пространства (X,ρ), в которых весь второй класс сводится только к одной метрике; иначе говоря, в таких пространствах есть только один способ построить трансляционно инвариантную метрику и, тем самым, все вышеперечисленные метрики на симплексе мер – Канторовича–Рубинштейна, Хаусдорфа–Куратовского и т.д. – совпадают! Это введенные в [8] и полностью описанные там *линейно жесткие* пространства, допускающие единственное невырожденное изометричное вложение в банахово пространство (невырожденность означает плотность линейной оболочки образа при вложении). Примером такого пространства служит универсальное пространство Урысона, что было впервые обнаружено Р. Холмсом [8], но это не единственный пример, см. [8]. Истинная причина этого эффекта опять-таки кроется в устройстве липшицевых функций. Оказывается, оно связано с тем, что любая липшицева функция с точностью до малого ϵ и некоторой константы представима в виде $f(z) = r(x,z)$ при некотором x. Эти эффекты изучены пока недостаточно.

Мы не обсуждаем здесь некоммутативные версии и обобщения метрик и их поднятий, которые в настоящее время активно рассматриваются (см., например, [11]).

В готовящейся к печати работе Ф. В. Петрова, П. Б. Затицкого и автора [3] введена следующая метрика, напоминающая метрику Леви–Прохорова; она определена в пространстве измеримых функций на квадрате $(X \times X, \mu \times \mu)$:

$$\rho(f, g) = \operatorname{dist}(f - g, 0);$$
$$\operatorname{dist}(h, 0) = \inf_{\substack{Y_{1,2} \subset X; \\ \mu Y_{1,2} > 1-\epsilon}} \{\epsilon : \sup_{x \in Y_1, y \in Y_2} h(x, y) < \epsilon\}.$$

Смысл этой метрики таков: расстояние между функциями измеряется малостью их разности, но не на произвольном подмножестве большой меры квадрата, как это требуется в метрике сходимости по мере, а на специальном подмножестве – прямоугольнике меры, произвольно близкой к единице. Она названа *метрикой раздельной сходимости по мере* и оказывается полезной для теорем вложения.

Ее аналог для мер выглядит так:

$$\operatorname{dist}(\mu_1, \mu_2) = \inf_{\substack{Y_{1,2} \subset X; \\ \mu Y_{1,2} > 1-\epsilon}} \{\epsilon : k_r(\mu_1|_{Y_1}, \mu_2|_{Y_2}) < \epsilon\},$$

где $\mu_i|_{Y_i}$, $i = 1, 2$, – нормированное ограничение меры μ_i на Y_i, $i = 1, 2$, а k_r – метрика Канторовича.

Литература

1. В. И. Богачев, А. В. Колесников, *Задача Монжа–Канторовича: достижения, связи и перспективы.* — Успехи мат. наук **67**, вып. 5 (2012), 3–110.
2. А. М. Вершик, *Метрика Канторовича: начальная история и малоизвестные применения.* — Зап. научн. семин. ПОМИ **312** (2004), 69–85.
3. А. М. Вершик, П. Б. Затицкий, Ф. В. Петров, *Виртуально непрерывные измеримые функции многих переменных и теоремы вложения*, готовится к печати.
4. Л. В. Канторович, *Математические методы организации и планирования производства.* Л., Изд–во ЛГУ, 1939.
5. Л. В. Канторович, *О перемещении масс.* — Докл. АН СССР **37**, вып. 7–8 (1942), 227–229.
6. Л. В. Канторович, *Об одной проблеме Монжа.* — Успехи мат. наук **3**, вып. 2 (1948), 225–226.
7. Л. В. Канторович, Г. Ш. Рубинштейн, *Об одном функциональном пространстве и некоторых экстремальных задачах.* — Докл. АН СССР **115**, вып. 6 (1958), 52–59.
8. R. Holmes, *The universal separable metric space of Urysohn and isometric embeddings thereof in Banach spaces.* — Fund. Math. **140**, no. 3 (1992), 199–223.
9. J. Melleray, F. Petrov, A. Vershik, *Linearly rigid metric spaces and the embedding problem.* — Fund. Math. **199**, no. 2 (2008), 177–194.

10. F. Otto, *The geometry of dissipative evolution equations: the porous medium equation.* — Comm. Partial Diff. Eqs. **26**, no. 1–2 (2001), 101–174.
11. M. Rieffel, H. Li, *Metric aspects of noncommutative homogeneous spaces.* — J. Funct. Anal. **257**, no. 7 (2009), 2325–2350.
12. V. Strassen, *The existence of probability measures with given marginals.* — Ann. Math. Stat. **36**, no. 2 (1965), 423–439.
13. A. Vershik, *Polymorphisms, Markov processes, and quasi-similarity.* — Discrete Contin. Dyn. Syst. **13**, no. 5 (2005), 1305–1324.
14. A. Vershik, *Long history of the Monge–Kantorovich transportation problem.* — Math. Intelligencer **35**, no. 4 (2013).
15. C. Villani, *Optimal Transport, Old and New.* Springer, 2006.

Vershik A. M. Two ways to define compatible metrics on the simplex of measures.

We introduce two general methods of lifting a metric on a space to the simplex of probability measures on the metric space. The first one is the method of transportation plans, or the coupling method; the second one is the method of considering norms dual to the restrictions of the Lipschitz norm to subspaces. The intersection of these two classes of metrics consists of the Kantorovich metric.

С.-Петербургское отделение
Математического института
им. В. А. Стеклова РАН,
наб. р. Фонтанки, д. 27,
191023 С.-Петербург, Россия
E-mail: avershik@gmail.com

Поступило 22 февраля 2013 г.