**А. В. Колесников**

# Транспортная задача и концентрация

Первые результаты о концентрации были получены П. Леви в его книге по функциональному анализу [10]. Само название «концентрация мер» было предложено В. Мильманом. Благодаря ему же явление концентрации приобрело большую популярность в математическом сообществе и нашло многочисленные приложения в функциональном анализе, геометрии, вероятности, комбинаторике и технических науках.

Среди сугубо математических приложений упомянем: 1) новое доказательство теоремы Дворецкого о «почти круглых» сечениях выпуклых тел, 2) изопериметрические теоремы сравнения М. Громова для многообразий положительной кривизны Риччи, 3) приложения в теории гауссовских случайных процессов (например, оценки статистического супремума), 4) применения к другим функциональным и вероятностным неравенствам (неравенства типа Соболева, неравенства типа Брунна—Минковского для выпуклых тел и т.п.). Подробнее об этом можно узнать в книгах [2; 7–9]. О вероятностных приложениях см. статью [11]. Для ознакомления с недавними результатами о концентрации и функциональных неравенствах для логарифмически вогнутых распределений также рекомендуем статью [13]. Теоремы о концентрации также позволяют оценить скорость сходимости системы к равновесному состоянию (см. комментарий ниже и приложение Е. В. Гасниковой настоящего пособия).

Классический пример свойства концентрации дает стандартное нормальное (гауссовское) $d$-мерное распределение $\gamma$. Как известно, плотность такого распределения задается формулой

$$\rho_\gamma(x) = \frac{1}{(2\pi)^{d/2}} \exp\Big(-\frac{|x|^2}{2}\Big).$$

Для произвольного множества $A$ со свойством $\gamma(A) > \frac{1}{2}$ рассмотрим его $r$-окрестность:

$$A_r = \{x : \exists y \in A : |x-y| \le r\}.$$

Тогда выполнено следующее неравенство концентрации:

$$\gamma(A_r) \geq 1 - \frac{1}{2} e^{-\frac{r^2}{2}}. \tag{1}$$

Как мы видим, $P(A_r)$ очень быстро (квадратично экспоненциально) стремится к единице.

Равномерное распределение $\sigma$ на единичной сфере $S^{d-1} \subset R^d$ также обладает аналогичным свойством:

$$\sigma(A_r) \geq 1 - \left(\frac{\pi}{8}\right)^{\frac{1}{2}} e^{-d\frac{r^2}{2}}. \tag{2}$$

Одно из важных следствий неравенств такого типа — неравенства для колебаний липшицевых функций. Пусть $f$ — 1-липшицева функция, т.е. удовлетворяющая соотношению $|f(x) - f(y)| \leq |x - y|$. Используя формулу коплощади

$$\int g(x)\ d\gamma = \int_{-\infty}^{\infty} \gamma(\{g > t\})\ dt,$$

из (1) можно получить неравенство вида

$$\gamma\big(\{x : f(x) - \int f\ d\gamma > t\}\big) \leq 2e^{-ct^2}$$

для некоторой универсальной константы $c$. Полученное свойство обычно формулируется в виде «липшицевы функции с большой долей вероятности мало отличаются от своего среднего значения».

Несмотря на простой вид, доказательство (1) нетривиально. Классический способ основан на описании так называемых «изопериметрических множеств» — множеств, имеющих наименьшую границу среди множеств такой же меры (вероятности). В евклидовом пространстве таким, как известно, является шар. На сфере их роль выполняют сферические «шары», а в пространстве, наделенном гауссовым распределением, — полупространства $\{x : \langle x, a\rangle < c\}$, $a \in \mathbb{R}^d$, $c \in \mathbb{R}$. Это — хорошо известный в теории вероятностей результат, доказанный В. Судаковым и Б. Цирельсоном (и независимо от них К. Бореллем, см. [1]). Из того факта, что $r$-окрестности полупространств являются полупространствами (т.е. опять изопериметрическими множествами), несложно извлечь следствие, что функция

$r \mapsto F(A,r) = \gamma(A_r)$ среди всех множеств фиксированной вероятности $p$ растет медленнее всего для изопериметрического $A$. Для этого множества функция $F(A,r)$ явно вычисляется, и мы получаем (1).

Указанный способ плох тем, что явно найти изопериметрические множества в более общем случае невозможно. Существует несколько подходов к доказательству неравенства концентрации. В настоящем пособии мы опишем связь явления концентрации с транспортной задачей, возникшей и развившейся в совершенно другой области математики. Связь эта была найдена в работе К. Мартон [12].

Транспортная задача ведет свою историю от классической работы Г. Монжа [14], написанной в 1781 году. В этой работе задача была сформулирована следующим образом: имеется куча песка и яма одинаковых объемов. Как засыпать песком яму, потратив наименьшие усилия на перевозку? Конечно, это не единственная возможная «экономическая» формулировка транспортной задачи. Речь, например, может идти о перевозе грузов со складов по заданным адресам.

В дискретной постановке мы имеем набор точек $\{x_i\}$, $1 \le i \le N$. Задано $N$ других точек $\{y_i\}$ и фукция стоимости $c(x,y)$ (например, расстояние или квадрат расстояния). Как построить взаимно-однозначное отображение $T$, сопоставляющее каждой точке из первого набора точку из второго набора, так, чтобы суммарная стоимость $\sum_{i=1}^{N} c(x_i, y_i)$ была наименьшей?

В дальнейшем транспортная задача переживала как периоды забвения, так и бурного развития. На языке современной математики транспортная задача была переформулирована и решена Л. Канторовичем в 40-х годах XX-го века (см. [3]) и получила в дальнейшем название задачи Монжа—Канторовича. Важным шагом в работах Канторовича было применение развитого им в теории линейного программирования метода двойственности и формулировка транспортной задачи на языке теории меры и функционального анализа. О приложениях двойственности и задач типа транспортной в технических науках см., например, главу 2 и приложение Е. В. Гасниковой настоящего пособия.

Пусть задана пара вероятностных распределений $\mu$ и $\nu$ на пространствах $X = Y = \mathbb{R}^d$. Решением задачи Монжа—Канторовича называется распределение $m$ на $\mathbb{R}^{2d}$, удовлетворяющее следующим условиям:

1. Проекции $m$ на $X$ и $Y$ равны соответственно $\mu$ и $\nu$:

$$\mathrm{pr}_X m = \mu, \ \ \mathrm{pr}_Y m = \nu. \tag{3}$$

2. Распределение $m$ реализует минимум следующего функционала:

$$\mathcal{F}(m) : m \to \int_{X\times Y} c(x,y)\ dm,$$

где $c : X\times Y \to \mathbb{R}$ — некоторая функция, называемая функцией стоимости (cost function).

При весьма общих предположениях задача Монжа—Канторовича имеет решение.

В дальнейшем мы будем интересоваться только случаем $c(x,y) = |x-y|^2$.

Обратим теперь внимание на важное отличие задачи Монжа—Канторовича от исходной задачи Монжа. В задаче Монжа речь идет о *перевозе* груза, что на математическом языке соответствует задаче существования *отображения* $T : X \to Y$, преобразующего распределение $\mu$ в распределение $\nu$ (последнее означает, что $\nu(A) = \mu(\{x : T(x)\in A\})$ и реализующего минимум функционала

$$W_2^2(\mu,\nu) = \int_{\mathbb{R}^d} |x - T(x)|^2 d\mu.$$

Оказывается, что при весьма общих условиях (например, распределения $\mu$ и $\nu$ непрерывны) эти задачи эквивалентны. Если $m$ — решение задачи Монжа—Канторовича, то $m$ сосредоточено на графике некоторого отображения $T$: $m\{(x,y) : y = T(x)\} = 1$. Мы будем называть $T$ оптимальным отображением. Существование $T$ было доказано Я. Бренье в [5]. Более того, имеет место следующий удивительный факт.

**Теорема 1.** *$T$ имеет вид*

$$T(x) = \nabla\varphi(x),$$

*где $\varphi$ — некоторая выпуклая функция.*

Величина

$$W_2(\mu,\nu) = \sqrt{\int_{\mathbb{R}^d} |x - T(x)|^2 d\mu}$$

называется расстоянием Канторовича (также можно встретить названия «расстояние Канторовича—Рубинштейна» и «расстояние Васерштейна»). Действительно, можно проверить, что $W_2(\mu, \nu)$ является расстоянием на пространстве вероятностных распределений.

Пусть теперь распределения $\mu$ и $\nu$ заданы плотностями $\mu = \rho_1\, dx$, $\nu = \rho_2\, dx$. Свойство $T$ отображать $\mu$ в $\nu$ аналитически записывается с помощью формулы замены переменной:

$$\rho_2(\nabla\varphi)\det D^2\varphi = \rho_1.$$

Если рассматривать $\varphi$ как неизвестную функцию, то мы получаем *уравнение Монжа-Ампера.* Под $D^2\varphi = D(\nabla\varphi)$ подразумевается матрица вторых производных (гессиан) функции $\varphi$.

Сделаем важное техническое замечание. Выполнено очевидное тождество

$$\int_{\mathbb{R}^d} |x-y|^2\, dm = \int_{\mathbb{R}^d} |x|^2 d\mu - 2\int_{\mathbb{R}^d} \langle x, y\rangle dm + \int_{\mathbb{R}^d} |y|^2 d\nu.$$

Поэтому поиск минимума функционала $\int_{\mathbb{R}^d} |x-y|^2\, dm$ эквивалентен поиску максимума функционала $\int_{\mathbb{R}^d} \langle x, y\rangle\, dm$.

Существование $\varphi$ может быть доказано разными способами. Стандартный подход состоит в применении метода двойственности Канторовича и работе с так называемыми циклически монотонными множествами. При этом выпуклость $\varphi$ получается автоматически. Двойственная задача Канторовича принимает вид

$$\int \Phi(x)\, d\mu + \int \Psi(y)\, d\nu \to \max,$$

где функционал максимизируется среди функций, удовлетворяющих условию $\Phi(x) + \Psi(y) \le |x-y|^2$. Отображение $T$ связано с $\Phi$ следующим образом: $T(x) = x - \nabla\Phi(x)$ (см. подробнее гл. 1, [16]).

Формальное, но поучительное доказательство того факта, что $T$ является градиентом, можно получить путем вывода уравнения Эйлера—Лагранжа (см. [6]). Пусть $T$ — произвольное отображение из $\mu$ в $\nu$. Решение задачи Монжа—Канторовича можно искать, как условный экстремум функционала

$$\int_{\mathbb{R}^d} \langle T(x), x\rangle d\mu$$

при условии $\rho_\nu(T)\det DT = \rho_\mu$. Составим функционал Лагранжа:

$$\int_{\mathbb{R}^d} \Big( \langle T(x), x\rangle \rho_\mu + \lambda(x)(\rho_\nu(T)\det DT - \rho_\mu) \Big) dx.$$

Функция $\lambda$ играет роль множителя Лагранжа. Чтобы найти первую вариацию функционала Лагранжа, рассмотрим инфинитезимальную вариацию

$$T_\varepsilon(x) = T(x) + \varepsilon\omega(x)$$

отображения $T$. Здесь $\omega$ — гладкое векторное поле с компактным носителем. Очевидно,

$$\rho_\nu(T_\varepsilon) \approx \rho_\nu(T) + \varepsilon\langle \omega, \nabla\rho_\nu(T)\rangle.$$

Можно проверить, что

$$\det(DT + \varepsilon D\omega) = \det DT \cdot \det(I + \varepsilon(DT)^{-1}D\omega) \approx$$
$$\approx \det DT \Big( 1 + \varepsilon \mathrm{Tr}\big[DT^{-1}\cdot D\omega\big] \Big).$$

Таким образом, первая вариация функционала Лагранжа равна

$$\int_{\mathbb{R}^d} \Big( \langle \omega(x), x\rangle \rho_\mu + \lambda(x)\cdot \rho_\nu(T)\mathrm{Tr}\big[DT^{-1}D\omega\big] +$$
$$+ \lambda(x)\langle \nabla\rho_\nu(T), \omega\rangle \, \frac{\rho_\mu}{\rho_\nu(T)} \Big) dx.$$

Заметим, что $\mathrm{div}(\omega(T^{-1})) = \mathrm{Tr}D\big[\omega(T^{-1})\big] = \mathrm{Tr}\big[DT^{-1}D\omega\big](T^{-1})$.

Интегрируя по частям и применяя замену переменных, несложно убедиться в том, что

$$\int_{\mathbb{R}^d} \lambda(x)\cdot \mathrm{Tr}\big[DT^{-1}\cdot D\omega\big]\rho_\mu \, dx = \int_{\mathbb{R}^d} \lambda(T^{-1})\mathrm{div}(\omega(T^{-1}))\rho_\nu \, dx =$$
$$= -\int_{\mathbb{R}^d} \langle \nabla\big[\lambda(T^{-1})\big], \omega(T^{-1})\rangle \rho_\nu \, dx - \int_{\mathbb{R}^d} \lambda(T^{-1})\big\langle \omega(T^{-1})), \frac{\nabla\rho_\nu}{\rho_\nu}\big\rangle \rho_\nu \, dx.$$

Следовательно, вариация равна

$$\int_{\mathbb{R}^d} \Big( \langle \omega(x), x\rangle \rho_\mu - \int_{\mathbb{R}^d} \langle \nabla\big[\lambda(T^{-1})\big], \omega(T^{-1})\rangle \rho_\nu \, dx = 0.$$

Пусть $\lambda = u(T)$, где $u$ — некоторая функция. Применяя опять формулу замены переменных, получаем, что для любого гладкого поля $\omega$ выполнено

$$\int_{\mathbb{R}^d} \Big( \langle \omega(x), x \rangle - \langle \nabla u(T), \omega(x) \rangle \Big) \rho_\mu dx = 0.$$

Следовательно:

$$\nabla u(T) = x \implies T^{-1} = \nabla u.$$

Таким образом, $T^{-1} = \nabla u$. В силу симметричности задачи относительно $\mu$ и $\nu$ то же утверждение можно сделать для самого отображения $T$.

В качестве иллюстрации эффективного использования оптимальной транспортировке в анализе приведем доказательство М. Громова классического изопериметрического неравенства.

**Пример 1.** *Среди множеств фиксированной меры Лебега шары имеют наименьшую поверхностную меру.*

**Доказательство.** Пусть $A \subset \mathbb{R}^d$ — борелевское множество, $B_r = \{x : |x| \le r\}$ — шар, удовлетворяющий условию $\lambda(A) = \lambda(B_r)$, где $\lambda$ — мера Лебега на $\mathbb{R}^d$. Пусть $T = \nabla\varphi$ — оптимальная транспортировка, отображающая $\lambda|_A$ в $\lambda|_{B_r}$. По формуле замены переменных $\det D^2\varphi = 1$ на $A$ (для простоты изложения считаем, что $\varphi$ — гладкая функция, хотя аргументы ниже легко обобщаются на негладкий случай). Матрица $D^2\varphi$ симметрична и неотрицательна, поэтому $1 = \sqrt[d]{\det D^2\varphi} \le \frac{\Delta\varphi}{d}$ в силу неравенства между средним арифметическим и средним геометрическим. Проинтегрируем это неравенство по $A$ и применим теорему Остроградского-Гаусса:

$$d\lambda(A) \le \int_A \Delta\varphi \, dx = \int_{\partial A} \big\langle \nabla\varphi, n_A \big\rangle \, d\mathcal{H}^{d-1} \le r\mathcal{H}^{d-1}(\partial A).$$

Здесь $n_A$ — единичная нормаль к $\partial A$, $\mathcal{H}^{d-1}$ — $(d-1)$-мерная мера Хаусдорфа. Из соотношения $\lambda(A) = \lambda(B_r) = \frac{\pi^{\frac{d}{2}}}{\Gamma(1+\frac{d}{2})} r^d$ получаем классическое изопериметрическое неравенство

$$\lambda^{1-\frac{1}{d}}(A) \le \kappa_d \mathcal{H}^{d-1}(\partial A),$$

где $\kappa_d = \frac{\left[\Gamma(1+\frac{d}{2})\right]^{\frac{1}{d}}}{d\sqrt{\pi}}$. Из доказательства следует, что неравенства становятся равенствами в случае $A = \{\|x - x_0\| \le r\}$. Таким образом, шары имеют наименьшую поверхностную меру среди множеств фиксированной меры Лебега. $\square$

Пусть $\nu$ — некоторое вероятностное распределение. Энтропией вероятностного распределения $g \cdot \nu$ относительно $\nu$ называется величина $\mathrm{Ent}_\nu(g) = \int g \log g \; d\nu$ (мы считаем, что функция $x \log x$ равна нулю в точке 0).

Следующий результат, доказанный М. Талаграном [15], связывает теорию оптимальной транспортировки с функцинальными неравенствами.

**Теорема 2.** *Пусть $\mu = \gamma$ — стандартное гауссовское распределение. Предположим, что $\nu = g \cdot \gamma$ — другое вероятностное распределение. Тогда квадрат расстояния Канторовича между этими распределениями оценивается относительной энтропией $g$:*

$$\frac{1}{2} W^2(\gamma, g \cdot \gamma) \leq \int_{\mathbb{R}^d} g \cdot \log g \; d\gamma := \mathrm{Ent}_\gamma(g).$$

**Доказательство.** Для простоты изложения предположим, что $g$ и $\varphi$ — гладкие функции (это бывает не всегда, но общий случай можно свести к этому). Рассмотрим формулу замены переменной:

$$e^{-\frac{x^2}{2}} = g(\nabla\varphi) e^{-\frac{|\nabla\varphi|^2}{2}} \det D^2\varphi.$$

Прологарифмируем это соотношение:

$$-\frac{x^2}{2} = \log g(\nabla\varphi) - \frac{|\nabla\varphi|^2}{2} + \log \det D^2\varphi.$$

Перепишем его в виде

$$\frac{1}{2}|x - \nabla\varphi|^2 = \langle x, x - \nabla\varphi\rangle + \log g(\nabla\varphi) + \log \det D^2\varphi.$$

Проинтегрируем полученное неравенство по $\gamma$. Заметим, что $\nabla e^{-\frac{x^2}{2}} = -x e^{-\frac{x^2}{2}}$. Из формулы интегрирования по частям следует:

$$\int_{\mathbb{R}^d} \langle x, x - \nabla\varphi\rangle \; d\gamma = \int_{\mathbb{R}^d} \left(d - \mathrm{Tr} D^2\varphi\right) \; d\gamma$$

(напомним, что $d$ — размерность). Следовательно,

$$\frac{1}{2}\int_{\mathbb{R}^d} |x - \nabla\varphi|^2 d\gamma + \int_{\mathbb{R}^d} \left(\mathrm{Tr} D^2\varphi - d - \log \det D^2\varphi\right) d\gamma \leq \int_{\mathbb{R}^d} \log g(\nabla\varphi) d\gamma.$$

Заметим теперь, что

$$\mathrm{Tr} D^2\varphi - d - \log\det D^2\varphi \geq 0.$$

Действительно, если $\lambda_i$ — собственные значения матрицы $D^2\varphi$, то

$$\mathrm{Tr} D^2\varphi - d - \log\det D^2\varphi = \sum_{i=1}^{d} \lambda_i - 1 - \log\det\lambda_i \geq 0$$

(в силу неотрицательности функции $x - 1 - \ln x$). Таким образом, получаем

$$\frac{1}{2}\int_{\mathbb{R}^d} |x - \nabla\varphi|^2 d\gamma \leq \int_{\mathbb{R}^d} \log g(\nabla\varphi) d\gamma = \int_{\mathbb{R}^d} g \log g \ d\gamma.$$

□

**Теорема 3 (К. Мартон).** *Если вероятностное распределение $\mu$ удовлетворяет неравенству Талаграна:*

$$W^2(\mu, g\cdot\mu) \leq C\int_{\mathbb{R}^d} g\cdot\log g \ d\mu,$$

*то $\mu$ удовлетворяет неравенству гауссовской концентрации:*

$$\mu(A_r) \geq 1 - 2e^{-\frac{r^2}{4C}}, \ \mu(A) \geq \frac{1}{2}. \tag{4}$$

*В частности, неравенству гауссовской концентрации удовлетворяет гауссовское распределение.*

**Доказательство.** Положим $(A^r)^c = \mathbb{R}^d \backslash A^r$. Рассмотрим оптимальную транспортировку $\nabla\varphi$ вероятностного распределения $\mu_1 = \frac{1}{\mu(A)}\times$ $\times I_A\cdot\mu$ в вероятностное распределение $\mu_2 = \frac{1}{\mu((A^r)^c)} I_{(A^r)^c}\cdot\mu$. В силу того, что расстояние между носителями $\mu_1, \mu_2$ превосходит $r$, имеем $W_2(\mu_1,\mu_2) \geq r$. В силу неравенства треугольника (напомним, что $W_2$ — расстояние):

$$r \leq W_2(\mu_1,\mu_2) \leq W_2(\mu_1,\mu) + W_2(\mu_2,\mu).$$

По неравенству Талаграна

$$r \leq \sqrt{2C\mathrm{Ent}_\mu\mu_1} + \sqrt{2C\mathrm{Ent}_\mu\mu_2}.$$

Так как $\mathrm{Ent}_\mu \mu_1 = \log \frac{1}{\mu(A)}$, $\mathrm{Ent}_\mu \mu_2 = \log \frac{1}{\mu((A^r)^c)}$, немедленно получаем

$$\frac{r^2}{4C} \le \log \frac{1}{\mu(A)} + \log \frac{1}{\mu((A^r)^c)}.$$

Следовательно,

$$\mu((A^r)^c) \le 2e^{-\frac{r^2}{4C}}.$$

Остается заметить, что $\mu((A^r)^c) = 1 - \mu(A^r)$. Теорема доказана. $\square$

Несложно проверить, что приведенные выше аргументы применимы к случаю распределения с плотностью $e^{-V}$, где $D^2V \ge K \cdot \mathrm{Id}$, $K > 0$ (неравенство понимается в матричном смысле, эквивалентная формулировка: $\langle D^2V \cdot v, v\rangle \ge K$ для любого вектора $v \in \mathbb{R}^d$ единичной длины). В этом случае также получаем гауссовскую концентрацию. Те же самые аргументы работают для сферы или, более общим образом, для многообразия с ограниченным снизу тензором Риччи.

Напоследок кратко обсудим еще одно приложение транспортной задачи — оценку скорости сходимости к равновесному состоянию. Пусть $V$ — равномерно выпуклый потенциал:

$$D^2V \ge K \cdot \mathrm{Id}, K > 0.$$

Рассмотрим решение уравнения Фоккера—Планка:

$$\frac{\partial \mu_t}{\partial t} = \Delta \mu_t + \mathrm{div}(\nabla V \cdot \mu_t),$$

$\mu_t = \rho_t\, dx$ — вероятностное распределение. Оказывается, для двух решений этого уравнения выполнено неравенство

$$\frac{d}{dt} W_2^2(\mu_t, \nu_t) \le -2K \cdot W_2^2(\mu_t, \nu_t)$$

(см. [16], пример 9.10). Это можно проверить, непосредственно продифференцировав расстояние Канторовича по параметру $t$. Очевидно, эта оценка дает экспоненциальную скорость сходимости $\mu_t$ к равновесному распределению:

$$W_2(\mu_t, \nu_t) \le W_2(\mu_0, \nu_0) e^{-Kt}.$$

Приложения такого рода включают в себя широкий класс уравнений, являющихся градиентными потоками относительно метрики Канторовича. Подробнее об этом см. в [4].

# Литература


1. *Богачев В. И.* Гауссовские меры. М.: Наука, 1997.

2. *Зорич В. А.* Математический анализ задач естествознания. М.: МЦНМО, 2008.

3. *Канторович Л. В.* О перемещении масс // ДАН СССР. 1942. Т. 37. С. 227–229.

4. *Ambrosio L., Gigli N., Savaré G.* Gradient flows in metric spaces and in the Wasserstein spaces of probability measures. Lectures in Math. ETH Zurich, 2008.

5. *Brenier Y.* Polar factorization and monotone rearrangement of vector valued functions // Comm. Pure Appl. Math. 1991. V. 44. P. 375–417.

6. *Evans L. C.* Partial differential equations and Monge-Kantorovich mass transfer. In «Current developments in mathematics». Cambridge, 1997; Boston: Int. Press, 1999. P. 65–126. http://math.berkeley.edu/~evans/

7. *Gromov M.* Metric structure for Riemannian and non-Riemannian spaces. V. 152. Boston: Birkhäuser, 1998.

8. *Milman V., Schechtman G.* Asymptotic theory of finite dimensional normed vector spaces. Lect Notes in Math. V. 1200. Springer, 1986.

9. *Ledoux M.* The concentration of measure phenomenon. Mathematical Surveys and Monographs 89. Amer. Math. Soc., 2001.

10. *Levy P.* Problème concretes d'analyse fonctionelle. Paris: Gauthier-Villars, 1951.

11. *Lugosi G.* Concentration of measures inequalities. Barcelona, 2009. http://www.econ.upf.edu/~lugosi/anu.pdf

12. *Marton K.* A measure concentration inequality for contractive Markov chains // Geom. Func. Anal. 1997. V. 6. P. 556–571.

13. *Milman E.* On the role of Convexity in Isoperimetry, Spectral-Gap and Concentration // Invent. Math. 2009. V. 177. N. 1. P. 1–43.

14. *Monge G.* Mémoire sur la théorie des dédlais et de remblais. Histoire de l'Acad'emic Royale des science, 1781.

15. *Talagrand M.* Transportation cost for Gaussian and other product measures // Geom. Funct. Anal. 1996. V. 6. P. 587–600.

16. *Villani C.* Topics in Optimal Transportation. Amer. Math. Soc. Providence. Rhode Island, 2003.
http://math.univ-lyon1.fr/homes-www/villani/