

数学通报: 2013 年第 52 卷第 8、9 期

双边 Hardy 不等式及其几何应用*

陈木法

(北京师范大学)

(2013 年 6 月 11 日)

2014 年 7 月 8 日



今天要谈的是双边的 Hardy 不等式及其在几何中的应用。双边的 Hardy 不等式可能很多人不是特别熟悉，我会慢慢介绍。事实上，Hardy-type inequality 是调和分析中很大的一个主题。

下面分三个小标题，首先是背景介绍，第二个是今天的主题 Hardy-type inequality (Hardy-type 不等式)，第三个是几何应用，介绍的主要是这两年的工作。

*本文原刊于《数学传播》(台湾)2013 年第 37 卷第 2 期。感谢陈木法教授及《数学传播》同意本刊刊载他于 2012 年 10 月 22 日在台湾大学天文数学馆醉月湖讲座(由中央研究院数学研究所与台湾大学数学系联合主办)所做的演讲记录(整理者: 陈丽伍)。本文下方注解为《数学传播》所加，以便读者阅读。文中最后一图及其说明是此稿新增的。

§1 背景介绍

大概在上世纪,七十年代 (1970s), 也就是四、五十年前, 数学重新回归自然。一个代表性的例子就是概率论跟统计力学的交叉形成的新的研究领域。当年, 我们就是走这个被称为交互作用的粒子系统 (interacting particle system) 的研究方向。统计力学研究的中心问题当然是相变现象 (Phase Transition)。相变现象的数学是无穷维的数学。无穷维数学的工具很少, 所以要四处找工具。因为这个原因, 我寻访过计算数学 (获知其中的三对角线矩阵算法), 也寻访过泛函分析、几何代数、数学物理方法等等, 发现几何做的比较精彩。几何里面做了许多估计的工作, 所以我们就去学几何。随后不久, 发现概率方法也能证明几何结果, 这大概是在 1992 年, 我们开始做一点几何的估计。然后大约在 1999 年, 距今十多年了, 有个很大的发现, 发现调和分析, 也就是 Hardy 不等式¹ 这个工具对我们很有用。当时的十个判别准则全是在知道 Hardy 不等式以后才展开的、才完成的。十多年过去了, 大概就这两年, 我们才想到用概率方法也能做 Hardy-type 不等式, 由此迎来了新的篇章。

概率 (又称为几率) 的一个特点就是跟其他领域比较起来相对年轻, 所以比较开放。一方面, 概率理论是在数学好多分支的抚育下成长起来的; 另一方面, 最近这些年来, 概率的一些思想也渗透到很多其它的数学分支。今天讲的就是一个例子, 是概率应用于其它数学分支的一种比较深入的结果。同时, 也希望让年轻的朋友体会一下数学的功夫。我觉得做数学很大的程度要靠功夫, 靠硬功夫。

§2 Hardy-type 不等式

谈到 Hardy-type 不等式, 当然要从老 Hardy 不等式开始。老 Hardy 不等式是九十年前, 也就是 1920 年提出的一个不等式。这个不等式非常的简练。

$$\int_0^\infty \left(\frac{1}{x} \int_0^x f \right)^p dx \leq \left(\frac{p}{p-1} \right)^p \int_0^\infty f^p, \quad f \geq 0, p > 1$$

这就是 Hardy 在 1920 年提出的不等式。这里及随后, Lebesgue 测度 dx

¹Godfrey Harold Hardy (1877-1947), English mathematician, best known for his achievement in number theory and mathematical analysis.

常略去不写。还有一种写法把 0 到 x 局部的积分, 写成 $Hf(x)$ 。这个 H 是为了纪念 Hardy 而通行的记号, 称为 **Hardy** 算子。利用算子 H , 上式可改写成

$$\int_0^\infty \left(\frac{1}{x} Hf(x) \right)^p dx \leq \left(\frac{p}{p-1} \right)^p \int_0^\infty f^p. \quad \boxed{\text{常数 } \frac{p}{p-1} \text{ 精确}}$$

还有一个更方便的解释: 把左边改成离散的求和, 0 到 x 的部分和的平均就是算术平均, 左边就是算术平均的 p 阶矩 (moment), 不等式成为 f 的算术平均的 p 阶矩小于等于 f 本身的 p 阶矩乘以常数 $[p/(p-1)]^p$, 这样就特别容易记住:

$$\|Af\|_{L^p(dx)} \leq \frac{p}{p-1} \|f\|_{L^p(dx)}.$$

Hardy 为什么要证明这个不等式呢? 他是为了寻找 Hilbert 的一个定理的初等证明。Hilbert 定理讲的是离散的级数, 证明一个 2 重级数 (Series) 的收敛性。Hilbert² 的证明是用傅立叶分析 (Fourier Analysis), Hardy 质疑一个初等的级数的收敛性为什么要用傅立叶分析, 嫌它太 “高级”, 认为应当有一个初等的证明。所以 Hardy 就引进这么一个不等式, 让这个 2 重级数的和用一个单重级数的和来控制, 这是引出这个不等式的原因。Hardy 这篇文章主要讨论的是离散级数的情况, 很奇怪的是他同时也把连续 (continuous) 的情况写出来, 但是没有给出证明。这是 1920 年的文章。如果 $p = 2$, Hardy 不等式就成为 Poincaré inequality, 这个不等式我们做得更久, 也更老一点, 一八八几年时就提出来了, 所以 Poincaré³ 不等式要比 Hardy 不等式更早一点。这个非常简单的不等式, 实际上牵扯到那个时代的三位顶尖的数学家, 我认为当时最前面的三位数学家就是这三位。所以很好玩, 这么简单的数学, 牵扯了这么多一流的数学家。

看到这么简单的不等式, 大家都会想要自己证它一下。我想对学数学的人来说这个很要紧, 因为数学是做出来的, 不是读出来的, 更不象看戏那样就能看会的。所以遇到什么问题, 就想自己证明它一下。在这里, 首先容易联想到 $L^p(\mu)$ 的性质, 这里 μ 是概率测度 (measure)。不等式的左方的被积函数是关于概率测度 $\frac{1}{x} \int_0^x$ 的一阶矩的 p 次方, 它受控于关

²David Hilbert (1862-1943), German mathematician, one of the most influential and universal mathematician of the 19th and 20th centuries.

³Jules Henri Poincaré (1854-1912), French mathematician, theoretical physicist, engineer and a philosopher of science, excelled in all fields of the discipline as it existed during his lifetime.

于此测度的 p 阶矩。 L^p 就是 p 阶矩，这个矩所对应的范数 (norm) 有一个单调上升 (monotone increasing) 的性质，当 p 大于零时， $L^p(\mu)$ 范数关于 p 单调上升。有了这一点概念，就可以给出证明。

$$\begin{aligned} \int_0^\infty \left(\frac{1}{x} \int_0^x f(y) dy \right)^p dx &\leq \int_0^\infty \left(\int_0^x f(y)^p dy \right) \frac{dx}{x} \\ &= \int_0^\infty f(y)^p dy \boxed{\int_y^\infty \frac{dx}{x}} \quad (\text{由 Fubini 定理}) \\ &= \infty \end{aligned}$$

倒数第二行中的 $1/x$ 积出来是无穷大，所以这个结论没有用。我们的第一轮证明失败了。

我们不必为此感到灰心丧气，实际上这个不等式的证明也不是平凡的。之前提到过，Hardy 在 1920 年写出这个不等式但是没有给证明，这个不等式的证明是五年 (1925 年) 之后才发表的。所以其证明并不是平凡的，不能一下子就看出来。让我们回头看看在哪里摔的跤？这个证明问题就在 $1/x$ 是个发散积分 (divergent integral)。把 $1/x$ 修改一下，改成 $\frac{1}{x^{1+\delta}}$ ，就会变成一个收敛积分 (convergent integral)。这样，我们可将上面的论证重写如下：

$$\begin{aligned} \int_0^\infty \left(\frac{1}{x} \int_0^x f(y) dy \right)^p \frac{dx}{x^\delta} &\leq \int_0^\infty \left(\int_0^x f(y)^p dy \right) \frac{dx}{x^{1+\delta}} \\ &= \int_0^\infty f(y)^p dy \boxed{\int_y^\infty \frac{dx}{x^{1+\delta}}} \quad (\text{由 Fubini 定理}) \\ &= \frac{1}{\delta} \int_0^\infty f(y)^p \frac{dy}{y^\delta}, \quad \boxed{\forall \delta > 0, f \geq 0}. \end{aligned}$$

这个改变很多时候可能就不平凡，这就是绝地逢生。别人看没路可走，我们却走出一条路来。可惜所得到的结果并不是我们所要的 Hardy 不等式。该怎么回到 Hardy 不等式呢？也就是需要将最后一行中的分母 y^δ 去除，就有希望了。移除分母最简单的方法就是变量替换 (variable substitution)，将不等式两边的 f 改写成 g ：

$$\text{设 } g(y) = f(y^{1+\gamma}) y^\gamma, \quad \gamma = \frac{\delta}{p-1}, \quad \delta = 1$$

这个一看就知道是变量替换公式。这一步可说是神来之笔，用了这一个步骤后，答案就呼之欲出，回到 Hardy 不等式了。实际上，这个 δ 是最

小化的解，等于 1，如果要知道答案，把 δ 代进去就完了，但这当中包含了很深刻的道理。这样，我们就完成了 Hardy 不等式的证明。

这个证明我自己觉得相当神奇。实际上 Hardy 1925 年的证明之后，经过了 40 年，这个证明才由 E.K. Godunova⁴ 找到。儘管他的文章在 1970 年，也就是五年后翻译成英文，在中研院数学所的图书馆可以找到这篇文章的资料，但是实际上没有人注意到这个工作。之后又过了差不多 40 年的时间，在 2002 年 Stein Kaijser⁵ , Lars-Erik Persson⁶ 和 Anders Öberg⁷ 发表了一篇文章⁸，重新发表这个证明，这时候大家才注意到有一个很简单的方法来证明这个不等式，Hardy 始终不知道有这样一个证明。

这是一个满漂亮的证明，可以体现我第一个想说的话，体现出数学的功夫。有功夫能看出来，没有功夫还是不行。当然，这个 Hardy 不等式太简单了。说老实话，我在做这个方向的早期就知道，但是它对我来说没有什么用处，因为太简单了。如果熟悉随机过程，看看其对应的算子就知道这太简单了，只是一种非常特别的情况，没有太多用处，所以就要往前走。接下来的故事就满长的，因为要一步一步发展。

刚才的 Hardy 算子写法是这个样子，

$$\int_0^\infty \left(\frac{1}{x} Hf(x) \right)^p dx \leq \left(\frac{p}{p-1} \right)^p \int_0^\infty f^p \quad \text{或等价地}$$

$$\|Hf\|_{L^p(x^{-p}dx)} \leq \frac{p}{p-1} \|f\|_{L^p(dx)}$$

现在用另一个观点来看，左方老 Hardy 处理的是 Hf 的 L^p 范数 (norm)，针对的是测度 $x^{-p}dx$ 。当然，上式左右的两个测度都很简单，因为简单，所以下一个任务就要推广一下。把两边的测度换成一般的、抽象的 Borel 测度 $(x^{-p}dx, dx) \rightarrow (\mu, \nu)$ ，就变成非常一般的形式：

$$\|Hf\|_{L^p(\mu)} \leq A \|f\|_{L^p(\nu)}$$

⁴E.K. Godunova, Russian mathematician,

⁵Sten Kaijser, working in Uppsala University, Sweden

⁶Lars-Erik Persson, working in Lulea University of Technology, Sweden

⁷Anders Öberg, working in University College of Gävle, Sweden

⁸Kaijser, Sten; Persson, Lars-Erik; Öberg, Anders, On Carleman and Knopp's inequalities. J. Approx. Theory 117 (2002), no. 1, 140 - 151.

这就是今天所要讲的 **Hardy-type** 不等式。从现在开始, 为节省记号, 不等式中的常数 A 均指最佳常数。这个 Hardy-type 不等式经历了大约五十年, 约在 1970 年前后, 这个不等式才研究得比较完整, 有多篇文章, G. Talenti (1969), G. Tomaselli (1969), R.S. Chisholm & W.N. Everitt (1970-71) 等, 其中 B. Muckenhoupt 在 1972 年发表的文章⁹ 被引用得多一点, 时间上也晚一点。其他的文章, 例如 M. Artola (1968-69) 和 D.W. Boyd & J.A. Erdős (1972) 根本没有发表, 因为知道别人已经写出来了, 所以即使投稿了的稿件也自己撤回。总之, 这里已经牵扯到六篇文章。

下一步想想能不能再推广到更一般的 p 和 q :

$$\|Hf\|_{L^q(\mu)} \leq A\|f\|_{L^p(\nu)},$$

这个时候两边几乎对等了。但一边是 L^p , 另一边是 L^q , 难度又增加了。差不多又做了 20 年, 经过 P. Gurka (1984), E.N. Batuev & V.D. Stepanov (1989), 直到 1990 年 B. Opic & A. Kufner 才做得比较完整。应当说, 这种 (p, q) 情形也很早就注意到了, Hardy 本人在 1930 年就开始做一些特殊的情况 (G.H. Hardy & J.E. Littlewood, 1930), 不过这篇也没有发表。最后指出, 《Hardy-type Inequalities》¹⁰ 是一本专著, 总结了之前的故事, 是关于 Hardy 不等式比较经典的著作。

再往下说之前, 还要提一下, 总体的目标是用测度 μ 和 ν 来算出 A 的一个较好的估计 (当然, 如能 sharp 就更好)。回忆 Hf 是 f 从 0 到 x 的积分。如果将 Hf 改写成 f , 原本的 f 就成为 f' 。这样改写后, 左边就成为 f 的 q 阶范数, 右边就变成 f' 的 p 阶范数, 再将区间改成更广一些的从 $-M$ 到 N ($M, N \leq \infty$) 的区间, 我们得到如下形式的 Hardy-type 不等式:

$$\left(\int_{-M}^N |f|^q d\mu \right)^{\frac{1}{q}} \leq A \left(\int_{-M}^N |f'|^p d\nu \right)^{\frac{1}{p}}.$$

当然, 这个不等式也可以写成

$$\|f\|_{\mu, q} \leq A\|f'\|_{\nu, p}.$$

⁹Muckenhoupt, Benjamin, Weighted norm inequalities for the Hardy maximal function. Trans. Amer. Math. Soc. 165 (1972), 207 - 226.

¹⁰B. Opic and A. Kufner, Hardy-type Inequalities, Pitman Research Notes in Mathematics Series, Longman, New York, 1990

这样一写，就要小心。原来的 Hardy 算子初值为 0，是 0 到 x 的积分。所以新的函数 f 在左端点要等于 0，边界条件要写上： $f(-M) = 0$ 。事实上，这时候可以写两个边界条件： $f(-M) = 0$ 和 $f(N) = 0$ 。两个边界条件就是今天的主题，双边的 **Hardy-type** 不等式。在双边的時候，不能写成单边的 Hardy 算子的形式，尤其你由一个 0 初值出去，只要非负、非零，积分后就不可能再回到 0，所以不可能写成 Hardy 算子的形式。这样，前面的单边情形变换 $(Hf, f) \rightarrow (f, f')$ 前后没有差别，但是后面的双边情形就有不同。后者是今天讨论的主要情况。

因为这个题目很多人不熟悉，所以我要强调一下。这是相当时髦的题目，就我所知，总共已有五本书讨论这个题目。

- [1] Opic, B. and Kufner, A. *Hardy-type Inequalities*. Longman, New York, 1990.
- [2] Kufner, A. and Persson, L.E. *Weighted Inequalities of Hardy-type*. World Scientific, 2003.
- [3] Kufner, A., Maligranda, L. and Persson, L.E. *The Hardy Inequality: About its History and Some Related Results*. Vydavatel'sky Servis, 2007.
- [4] Kokilashvili, V., Meshki, A. and Persson, L.E. *Weighted Norm Inequalities for Integral Transforms with Product Weights*. Nova Sci. Publ., New York, 2010.
- [5] Maz'ya, V. *Sobolev Spaces with Applications to Elliptic Partial Differential Equations* (2nd Ed.). Springer, 2011.

第一本书是刚才提到的 1990 年的名著。之后的四本都是最近十年出现的著作。前三本书名中都有 Hardy 不等式或是 Hardy-type 不等式。第三本书谈的是历史，内容是好玩的。一个分支的研究状况，如果东西少，就没有历史可谈；这里边讲了许多历史，说明这个题目已经有了充分的发展。第四本书名中没有提到 Hardy 不等式，但是 weighted norm inequalities 完全是 Hardy 不等式，只不过做的是高维的 Hardy 不等式。第五本书标题也没有 Hardy 不等式，因为这本书有 894 页，Hardy 不等式只是其中一部分。

一开始的时候，我提过想要用谱理论、特征值理论来研究相变现象。我做了整整十年之后，还不知道有 Hardy-type 不等式，我只知道有古典

的 Hardy 不等式，而不知道有这么一般的 Hardy-type 不等式。这个一般的不等式对我来说是个非常大的 favor。知道这个就不得了，好多东西都做出来了，对我们帮助非常大。比如说大家可能听说过的 log-Sobolev 不等式。Log-Sobolev 不等式我们也做了几十年，但是其 criteria 只是因为有了 Hardy-type 不等式后才找到的。很多很要紧的结果就是因为用了 Hardy-type 不等式，所以我们十年中不断地用。我有一个列表，一维情形有十个 criteria。在当时被称为十大准则。Hardy-type 不等式为什么在调和分析中有如此地位就是因为它是很要紧的，核心的一个问题。然后大约在二年前 (2010 年)，我们找到反向的、原来的 Hardy 理论中没有的东西，引导我们往这个方向发展。

在讲我的结果之前，先解释一下记号。第一个是测度 $\hat{\nu}$ ，它是由测度 ν 的绝对连续部分 (absolutely continuous part) 导出的。接下来所有的事情都要通过测度 μ 和 $\hat{\nu}$ 表示出来，所以 $\hat{\nu}$ 是很要紧的。

$$\boxed{\hat{\nu}(dx) = \hat{\nu}_p(dx) = \left(\frac{d\nu^\#}{dx} \right)^{\frac{-1}{p-1}} dx} \quad \nu^\# : \nu \text{ 的绝对连续部分,}$$

此处需要假定 $\nu^\#$ 的密度非零，否则可用极限过渡。接着需要常数 $k_{q,p}$:

$$\boxed{k_{q,p} = \left(1 + \frac{q}{p'}\right)^{\frac{1}{q}} \left(1 + \frac{p'}{q}\right)^{\frac{1}{p'}}} \quad \boxed{\leq 2 \text{ 若 } q \geq p},$$

其中 p' 是 p 的共轭指数: $1/p + 1/p' = 1$ 。后面要用到的这个 $k_{q,p}$ ，当 $q = p$ 时是精确的；当 $q \neq p$ 时，可换成更小一些的精确常数，此处略去。这个式子不需要记住，只要知道 $k_{q,p}$ 永远小于等于 2。所以粗糙地讲，把 $k_{q,p}$ 当成 2 好了。有了这两个记号后，就可以进入今天的第一个主要结果。

定理 2.1 (陈, Acta Math. Sin. Eng. Ser. 2013). 在 Hardy-type 不等式中的最佳常数 A 满足

$$(1) \quad \boxed{A \leq k_{q,p} B^*} \text{ 如 } 1 < p \leq q < \infty \text{ 且 } \mu_{pp} = 0,$$

(2) $\boxed{A \geq B_*}$ 如 $1 < p, q < \infty$, 其中

$$B^* = \sup_{x \leq y} \frac{\mu[x, y]^{\frac{1}{q}}}{\left\{ \hat{\nu}[-M, x]^{\frac{q(1-p)}{p}} + \hat{\nu}[y, N]^{\frac{q(1-p)}{p}} \right\}^{\frac{1}{q}}},$$

$$B_* = \sup_{x \leq y} \frac{\mu[x, y]^{\frac{1}{q}}}{\left\{ \hat{\nu}[-M, x]^{1-p} + \hat{\nu}[y, N]^{1-p} \right\}^{\frac{1}{q}}},$$

此外, 当 $q \geq p$ 时, 我们有 $\boxed{B_* \leq B^* \leq 2^{\frac{1}{p} - \frac{1}{q}} B_*}$.

换言之, 最佳常数 A 有上、下控制, 分别由 (1) 和 (2) 给出. B^* 和 B_* 这两个表达式由测度 μ 和 $\hat{\nu}$ 完全刻画. 只是对于上界, 要一些条件: q 要大于等于 p , 测度 μ 要连续. 熟知测度 μ 的分解有三部分: 第一部分绝对连续, 第二部分奇异连续 (singularly continuous), 第三部分是纯点的 (discrete), 这里假定纯点是 0, 也就是 $\mu_{pp} = 0$, 没有纯跳的部分. 我们着重指出: 上界估计是最重要、最有用的东西. 如果把 (1) 中的条件都去掉, 会是什么情况, 目前还没有答案. 所以这里是一个开放的地盘, 还有很多的问题.

讲一下这个定理有那些漂亮的地方. 首先, 第一个漂亮的地方是上、下界相差的常数是 universal, 是一个普适常数, 跟测度没有关系. 第二个优点是这里有两个边界, 一个是 $-M$, 另一个是 N . M 和 N 都可以是无穷大. 那么边界条件 $f(-M) = 0$, 如果 M 是无穷大, 表示趋于负无穷大时 f 趋于 0. 两个边界, 一个 $-M$, 一个 N , 这两个边界在这两个常数中是对称的, 这是一个漂亮的地方. 无论是 B^* 或是 B_* , 这两个边界都是对称的, 这是第二个优点. 第三个优点, 如果不喜欢一个上、一个下, 只利用一个也是可以. 因为 B^* 可以由 B_* 统起来, 相差一个常数倍而已. 将这个常数倍放大, 写到第一行里, 这个新常数仍然是小于等于 2. 如果 $q = p$, 两个常数就变成一个,

$$B^* = B_* = \sup_{x \leq y} \frac{\mu[x, y]^{\frac{1}{q}}}{\left\{ \hat{\nu}[-M, x]^{1-p} + \hat{\nu}[y, N]^{1-p} \right\}^{\frac{1}{q}}},$$

这又是另一个优点. 最后一个优点, 我做的这个双边的, 可以回到单边. 如果把第二加项 $\hat{\nu}[y, N]$ 去掉, 那么 y 不起作用, 可取为 N , 因此 N 的边界条件就不起作用, y 就可以移到 N 那边, 就变成单边. 所以右边的

边界条件不起作用, 就变成单个边界条件, 这个表达式就变得非常简单。因为那项去掉后, 分母可以倒过来, 就很简单。如果把第一项 $\hat{\nu}[-M, x]$ 去掉, 因为是对称的, 边界条件就变成在右边。也就是

去掉 $\hat{\nu}[y, N]$, 得到 $\boxed{f(-M) = 0}$, $B^- = \sup_x \mu[x, N]^{\frac{1}{q}} \hat{\nu}[-M, x]^{\frac{p-1}{p}}$.

去掉 $\hat{\nu}[-M, x]$, 得到 $\boxed{f(N) = 0}$, $B^+ = \sup_y \mu[-M, y]^{\frac{1}{q}} \hat{\nu}[y, N]^{\frac{p-1}{p}}$.

总结起来, 我们所研究的 **Hardy-type** 不等式是

$$\text{在 } (-M, N) \text{ 上, } \|f\|_{\mu, q} \leq A \|f'\|_{\nu, p},$$

其中 μ 和 ν 是 Borel 测度。对于单边情形, 如 $f(-M) = 0$, 即左边 $= 0$, 最佳常数写成 A^- ; 如果右边 $= 0$, 最佳常数写成 A^+ 。区分开来, 这时候就有了以下这个定理:

定理 2.2. 令 $q \geq p$, 则 $B^\pm \leq A^\pm \leq k_{q,p} B^\pm$, 其中

$$B^- = \sup_x \mu[x, N]^{\frac{1}{q}} \hat{\nu}[-M, x]^{\frac{p-1}{p}},$$

$$B^+ = \sup_y \mu[-M, y]^{\frac{1}{q}} \hat{\nu}[y, N]^{\frac{p-1}{p}}.$$

这就回答前面说的, 从 1920 年到 1990 年间 70 年研究的结果, 就是这个结果。容易看出, 从后一定理到前一个, 有相当的跨度。

现在的问题是, 该如何从单边走向双边, 这是我的核心问题, 也是困难的地方。原本 $p = q = 2$ 时, 我们会证明。最早的发现, 也就是从这个地方发现的。只是那是在用了三大工具、经五个步骤后才做出来的, 所以很辛苦。但是我还是不甘心, 觉得这样简单的问题, 为什么需要那么大的工具, 特别是需要容度 (capacity) 等等。经过长时间的摸索, 才找到新的办法。下面要讲的就是这个新的证明。相信许多读者, 不论是经验老到的或是初学者, 都会在此处停下来, 想想能否找到自己的证明。

在介绍我们的证明之前, 让我们回顾一下, 单边的情形结论已经有了。现在, 我们把 $-M$ 跟 N 的区间中间加了一个点 θ , 将区间分成两部分。



令 A_{θ}^{-} 为在 $(-M, \theta)$ 上, 当 $f(-M) = 0$ 时的最佳常数,

令 A_{θ}^{+} 为在 (θ, N) 上, 当 $f(N) = 0$ 时的最佳常数.

加了这个点 θ 后, 变成两个单边。在左边这个半区间里关心的是左端点 $-M$, 在另一个半边关心的是右端点的 N , 然后将刚才的结果重写一次, 得到下面的结果。

令 $\boxed{q \geq p}$. 则我们有 $B_{\theta}^{\pm} \leq A_{\theta}^{\pm} \leq k_{q,p} B_{\theta}^{\pm}$, 其中

$$B_{\theta}^{-} = \sup_{x < \theta} \mu[x, \theta]^{\frac{1}{q}} \hat{\nu}[-M, x)^{\frac{p-1}{p}},$$

$$B_{\theta}^{+} = \sup_{y > \theta} \mu[\theta, y]^{\frac{1}{q}} \hat{\nu}[y, N]^{\frac{p-1}{p}}.$$

这个跟上面的差别, 从左边半区间看, 无非是用 θ 代表先前的 N ; 从右边半区间看, 无非是用 θ 代表先前的 $-M$; 也就是换一下记号而已。这是下面我们要用到的已知结果。

现在, 我开始证明我们的上界估计: $\boxed{k_{q,p} B^* \geq A}$. 首先, 改写 B^* :

$$\begin{aligned} B^* &= \sup_{x \leq y} \frac{\mu[x, y]^{\frac{1}{q}}}{\left\{ \hat{\nu}[-M, x)^{\frac{q(1-p)}{p}} + \hat{\nu}[y, N]^{\frac{q(1-p)}{p}} \right\}^{\frac{1}{q}}} \\ &= \left\{ \sup_{x \leq y} \frac{\mu[x, y]}{\hat{\nu}[-M, x)^{\frac{q(1-p)}{p}} + \hat{\nu}[y, N]^{\frac{q(1-p)}{p}}} \right\}^{\frac{1}{q}} \\ &=: \left\{ \sup_{x \leq y} \frac{\mu[x, y]}{\varphi(x) + \psi(y)} \right\}^{\frac{1}{q}} \end{aligned}$$

右方的分数可分拆, 然后使用合分比定理 (proportional property) 得出

$$\begin{aligned} \frac{\mu[x, y]}{\varphi(x) + \psi(y)} &= \frac{\mu[x, \theta] + \mu(\theta, y]}{\varphi(x) + \psi(y)} \quad \boxed{\theta \in [x, y]} \\ &\geq \frac{\mu[x, \theta]}{\varphi(x)} \bigwedge \frac{\mu(\theta, y]}{\psi(y)}, \end{aligned}$$

此处 $x \wedge y = \min\{x, y\}$, 类似地, $x \vee y = \max\{x, y\}$ 。两边对 θ 取上确界, 得

$$\frac{\mu[x, y]}{\varphi(x) + \psi(y)} \geq \sup_{\theta \in [x, y]} \left\{ \frac{\mu[x, \theta]}{\varphi(x)} \wedge \frac{\mu(\theta, y]}{\psi(y)} \right\}.$$

再对 $x \leq y$ 取上确界, 注意两个 $\sup$ 可交换, 得出

$$\begin{aligned} \sup_{x \leq y} \frac{\mu[x, y]}{\varphi(x) + \psi(y)} &\geq \sup_{x \leq y} \sup_{\theta \in [x, y]} \left\{ \frac{\mu[x, \theta]}{\varphi(x)} \wedge \frac{\mu(\theta, y]}{\psi(y)} \right\} \\ &= \sup_{\theta} \sup_{\theta \in [x, y]} \{ \dots \} \\ &= \sup_{\theta} \left\{ \left[\sup_{x \leq \theta} \frac{\mu[x, \theta]}{\varphi(x)} \right] \wedge \left[\sup_{y \geq \theta} \frac{\mu(\theta, y]}{\psi(y)} \right] \right\}. \end{aligned}$$

两边取 $1/q$ 次方, 完成了证明的第一步。

引理 2.3. $B^* \geq \sup_{\theta} (B_{\theta}^- \wedge B_{\theta}^+)$.

这里我们做了什么? 我们把两个单边的情况跟一个整体的双边的情况联系起来。然而, 右方只是最小值, 那是不够的, 我们需要的是最大值。由测度 μ 的连续性假定 $\mu_{pp} = 0$, 存在 $\bar{\theta}$ 使得 $B_{\bar{\theta}}^- = B_{\bar{\theta}}^+$, 当然更有

$$B^* \geq \sup_{\theta} (B_{\theta}^- \wedge B_{\theta}^+) \geq B_{\bar{\theta}}^-.$$

于是

$$\begin{aligned} k_{q,p} B^* &\geq (k_{q,p} B_{\bar{\theta}}^-) \vee (k_{q,p} B_{\bar{\theta}}^+) \\ &\geq A_{\bar{\theta}}^- \vee A_{\bar{\theta}}^+ \quad (\text{上面所述的已知定理}) \\ &\geq \inf_{\theta} (A_{\theta}^- \vee A_{\theta}^+) \\ &\geq A \quad (\text{分裂技术}). \quad \square \end{aligned}$$

最后一步是两个半区间上的最佳常数与整个区间上的最佳常数的比较, 是典型的分裂技术, 其证明也不长, 我们十多年前就会了, 此处不再解释。我们指出: 最后这一步及开头的定理 2.2, 都用到条件 $q \geq p$ 。

这个证明总共有三个要点, 第一个是分裂技术 (splitting technique); 第二个是 $\bar{\theta}$ 的使用, 做了一个过河拆桥的好事, 不需要将 $\bar{\theta}$ 解出, 过了河就拆了桥; 最后一个技巧就是使用合分比, 将上界估计证出来。这个证明做出来的时后, 真会让人高兴得跳起来, 因为是这么漂亮又这么简

单。所以这是一个让我们很得意的证明。实际上，当我说那个复杂的表达式很漂亮，可能很多人心里并不同意，因为还是有点复杂。我说它漂亮，是因为我喜欢，因为证明是这么漂亮，所以当然会喜欢。

我已经证明了双边的 Hardy 的情况，用概率的语言讲，这是 Transient 情形。还有 Ergodic 的情况。

$$\begin{aligned} \|f\|_{\mu,q} &\leq A \|f'\|_{\nu,p} && \text{Transient,} \\ \|f - \mu(f)\|_{\mu,q} &\leq A \|f'\|_{\nu,p} && \text{Ergodic.} \end{aligned}$$

如果 μ 是概率测度，自然考虑 Ergodic 情况。若只是将两边等于 0 的边界条件都去掉，那么所得不等式是平凡的，因为恒等于 1 的常数 f 给出 $A = \infty$ 。所以中间要减去一个概率测度的平均值。这是概率论最重要的情形。这种情形让我吃了许多苦头，之前提到的三大工具、五个步骤就是证明这种情形的平行的结果。然而，刚才我们所用证明的方法完全可以套过来，所以我们那个方法在发表的论文中至少用了三次。它不仅能处理这种情况，还能处理 Log-Sobolev 不等式 (Log-Sobolev inequality) 以及 Nash 不等式 (Nash inequality)。所以有好多的故事。

§3 几何应用

这最后十几分钟的时间我想跟大家讲应用，这个 Hardy-type 不等式有什么用处。现在给大家一个很简单情况的应用。如果 $q = p$ ，这些估计都能改进。如果 $q \neq p$ ，应该也可以改进，但是还没做出来，这是一个待解的问题 [校稿注：现已基本解决]。我们考虑如下的 $q = p = 2$ 的特殊例子：

$$\|f\|_{\mu,2} \leq A \|f'\|_{\nu,2} \quad \text{on } (-M, N), \quad f(N) = 0.$$

其中测度 μ 和 ν 来自微分算子：

$$\begin{aligned} L &= a(x) \frac{d^2}{dx^2} + b(x) \frac{d}{dx}, & C(x) &= \int_{\theta}^x \frac{b}{a}, \\ \boxed{\mu(dx)} &= \frac{e^C}{a}(x)dx, & \boxed{\nu(dx)} &= e^{C(x)}dx, \end{aligned}$$

$\theta \in (-M, N)$ 是一个参考点。有趣的是，最佳常数可写成

$$\boxed{A^2 = \lambda_0^{-1}},$$

其中 λ_0 是算子 L 的一个主特征值:

$$Lf = -\lambda_0 f, \quad f \neq 0, f(N) = 0.$$

此处, 当 $M < \infty$ 时, 还需补充边界条件 $f'(-M) = 0$. 这就建立了不等式与特征值之间的关系。一般地讲, 不等式是特征值的变分形式; 而特征值所呈现的特征方程是使不等式成为等式所导出的方程。所以刚才讲的所有关于不等式最佳常数的估计都是特征值的估计。在当前情况下, 有 μ 也有 ν , 我们真正用的不是 ν 而是 $\hat{\nu}$, $\hat{\nu}$ 在这里特别简单, $\hat{\nu}(dx) = e^{-C(x)}dx$ 。

我们有以下的一般结果:

定理 3.1. 对于给定的算子 L , 存在系列 $\{\underline{k}_n\}_{n \geq 1}$ 和 $\{\bar{k}_n\}_{n \geq 1}$, 使得当 $n \uparrow$ 时,

$$\uparrow \underline{k}_n^{-1} \leq \lambda_0 = A^{-2} \leq \bar{k}_n^{-1} \downarrow.$$

换言之, 我们可以构造出两串数列, 一串升, 一串降, 慢慢靠近 λ_0 。实际上, 我相信它们都收敛于 λ_0 , 但是还没有证出来。特别地, 我们写出当 $n = 1$ 时的下界和上界如下。

推论 3.2. 对于一般的二阶微分算子 L , 我们有 $\underline{k}^{-1} \leq \lambda_0 \leq \bar{k}^{-1}$, 其中

$$\begin{aligned} \underline{k} &= \sup_{x \in (-M, N)} \frac{1}{\sqrt{\psi(x)}} \int_{-M}^N \psi(\cdot \vee x) \sqrt{\psi} d\mu \\ \bar{k} &= \sup_{x \in (-M, N)} \frac{1}{\psi(x)} \int_{-M}^N \psi(\cdot \vee x) \psi d\mu \\ \psi(x) &= \hat{\nu}(x, N) = \int_x^N e^{-C}, \quad x \vee y = \max\{x, y\}. \end{aligned}$$

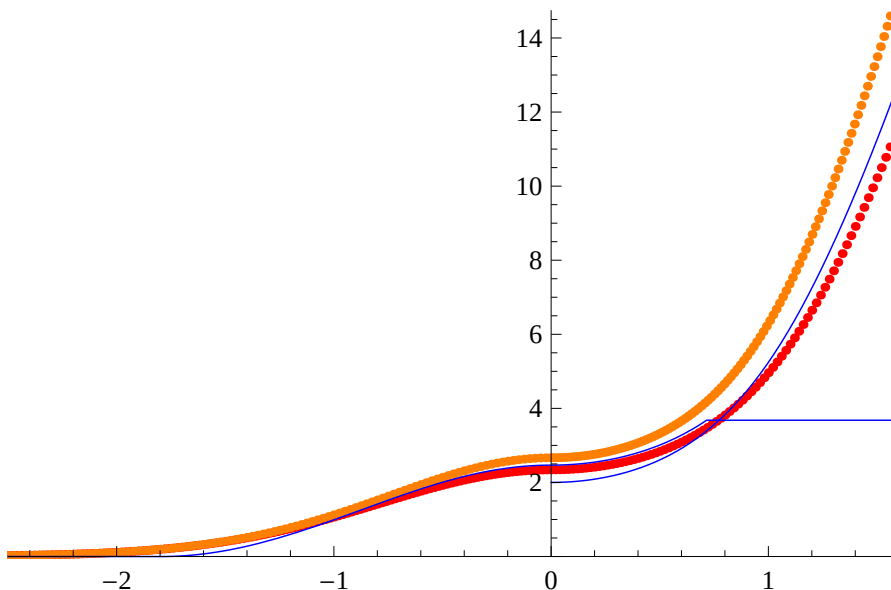
当然, 这里的上下界也是通过测度 μ 和 $\hat{\nu}$ 表出的, 适用于一般的算子。下面考虑更特殊的特例。

例 3.3. 令 $(-M, N) = (0, 1)$,

$$a(x) \equiv 1 \text{ 和 } b(x) = -(d-1)\alpha \tanh(\alpha x), \quad x \in (0, 1),$$

其中 d 和 α 为给定的常数参数。

将常数 d 和 α 代入 $\bar{k}$ 和 $\underline{k}$, 产生如下的图形。



画图时, 将常数 d 固定为 5, 看上下界随 α 的变动情况。因为 α 可能是纯虚数 (随后给出表达式), 所以这里我们做了变换: $\alpha = \sqrt{-\operatorname{sgn}(x)}|x|$, $x \in (-2.5, \pi/2)$. 图中有两条粗一点的曲线, 颜色深一点的是下界估计, 颜色浅一点的 (位于上方) 是上界估计。从图可以看出, 两条线相差最大的地方, 其倍数小于等于 2。左边几乎可以说是重叠, 这是画了图后才发现的。一般情形在做很多例子的时候, 上下界之比都小于等于 2, 比 4 要改进一倍, 但是没想到会如此贴近。这个贴近是很不寻常的发现, 因为负的时候更难, 允许无限, 在 -1 的地方就已经几乎相叠, 更别说更负的情形。

在这个图中除了两条粗一点的曲线外, 还有两条细一点的曲线。关于这两条细曲线的故事就很长了。上图的意义来自几何。考虑紧的黎曼流形 (Riemann Manifold), 假设无边, 因为是紧的, 所以谱值离散 (discrete spectrum), 考虑它的第一个非零的特征值 λ_1 。要点是用三个几何量做估计: 流形维数 d , 流形直径 D 和流形曲率的下界 K 。一个标准例子是 d 维单位球面:

$$\mathbb{S}^d : D = \pi, \operatorname{Ric} = d - 1, \lambda_1 = d \quad \forall d \geq 1.$$

为方便, 我们定义

$$\alpha = \frac{D}{2} \sqrt{\frac{-K}{d-1}} \quad \text{如 } K \neq 0, \quad \alpha = 0 \quad \text{如 } K = 0.$$

著名的 Myers 定理断言: 当 $K > 0$ 时, $|\alpha| \leq \pi/2$.

如开头所说,在几何中关于第一特征值有相当完整的研究。当 $K \geq 0$ 时,代表性成果如下表。前三个结果对于单位高维曲面达到最优,其中的第三个最好;第四个是在单位圆、零曲率时达到最优,被认为是几何中最艰深的结果。上一张图中的两条细曲线,在右方高一些的那一条就是下面的第三个结果。此处提及,这里我们只列出最优估计而不可能谈及其它结果,我们直奔目标而不顾细节。详细文献可在文末的论文中找到。

作者	下界估计: $K \geq 0$ 情形
A. Lichnerowicz (1958)	$\frac{d}{d-1} K$
P.H.Bérard, G.Besson & S. Gallot (1985)	$d \left\{ \frac{\int_0^{\pi/2} \cos^{d-1} t \, dt}{\int_0^{D/2} \cos^{d-1} t \, dt} \right\}^{2/d}, K = d - 1$
Chen & F.Y. Wang (1997)	$\frac{dK}{(d-1)(1 - \cos^d \alpha)}$
J.Q. Zhong & H.C. Yang (1984)	$\frac{\pi^2}{D^2}$

对于负曲率的情况,有三个最优估计。第二个优于第一个,它们都在 $\alpha = 0$ 处达到最优。第三个在曲率很负时优于第二个,但它们自身不能比较。

作者	下界估计: $K \leq 0$ 情形
H.C. Yang (1989) F. Jia (1991) $d \geq 5$ Chen & F.Y. Wang (1994)	$\frac{\pi^2}{D^2} e^{-(d-1)\alpha}$
Chen & F.Y. Wang (1997)	$\frac{1}{D^2} \sqrt{\pi^4 + 8(d-1)\alpha^2} \cosh^{1-d} \alpha$
Chen (1994)	$\frac{1}{D^2} ((d-1)\alpha \tanh \alpha \operatorname{sech} \theta)^2$

其中的 θ 可由如下方式得到: 定义

$$\begin{aligned} \theta_1 &= 2^{-1}(d-1)\alpha \tanh \alpha, \\ \theta_n &= \theta_1 \tanh \theta_{n-1}, \quad n \geq 2, \quad \text{则 } \theta_n \downarrow \theta. \end{aligned}$$

在零曲率的情况,最优估计如下表。其中第二个优于第一个。第三个在零附近优于第二个,误差可以达到 10^{-7} , 非常精确。

作者	下界估计: $K \approx 0$ 情形
11 authors	$\frac{\pi^2}{D^2} + \frac{K}{2}, \quad K \in \mathbb{R}$
Y.M. Shi & H.C. Zhang (2007)	$\sup_{s \in (0,1)} s \left[4(1-s) \frac{\pi^2}{D^2} + K \right] \quad (*)$
Chen & E. Scacciatelli & L. Yao (2002)	$\frac{\pi^2}{D^2} + \frac{K}{2} + (10 - \pi^2) \frac{K^2 D^2}{16}, \quad K \leq \frac{4}{D^2}$

其中的第二个结果有显式:

$$(*) = \begin{cases} \left(\frac{\pi}{D} + \frac{KD}{4\pi} \right)^2, & -4\pi^2 \leq KD^2 \leq 4\pi^2 \\ K, & KD^2 \in (4\pi^2, (d-1)\pi^2] \\ 0, & KD^2 < -4\pi^2 \end{cases}$$

总而言之, 从 1958 年到现在的这五、六十年间所得到的 10 个最优估计中, 经过比较, 以下 5 个结果领先。

作者	下界估计
Chen & F.Y. Wang (1997)	$\frac{dK}{(d-1)(1 - \cos^d \alpha)}, \quad K \geq 0$
Chen & F.Y. Wang (1997)	$\frac{1}{D^2} \sqrt{\pi^4 + 8(d-1)\alpha^2} \cosh^{1-d} \alpha$
Chen (1994)	$\frac{1}{D^2} ((d-1)\alpha \tanh \alpha \operatorname{sech} \theta)^2$
Y. Shi & H.C. Zhang (2007)	$\sup_{s \in (0,1)} s \left[4(1-s) \frac{\pi^2}{D^2} + K \right]$
Chen & E. Scacciatelli & L. Yao (2002)	$\frac{\pi^2}{D^2} + \frac{K}{2} + (10 - \pi^2) \frac{K^2 D^2}{16}, \quad K \leq \frac{4}{D^2}$

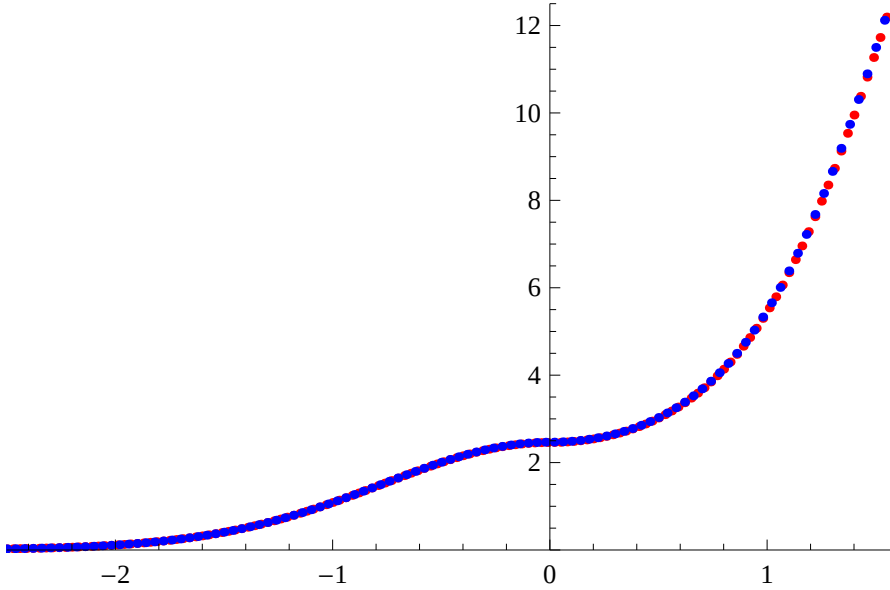
现在, 我们可以给出上图的完整的解释。我们给出了新的统一的下界估计, 它通常优于此表中的第二、第三个估计。在右端点近旁, 可用此表中的第一个估计作补充; 在零点近旁, 可用此表中第四个估计的改进形式作补充 (它即是图中的另一细曲线)。如前所述, 表中的第五个估计很精确, 但它只适用于零点的一个小邻域, 此处不再讨论。总之, 此图已显示出我们对于几何所获得的进步: 一个统一的下界加上两个补充。图中的上界估计最初的用意是作为判断下界估计优劣的参照物。然而, 图中显示出上、下界的曲线十分相似, 这就启发我们去做凸平均 (convex

mean)。挑选凸平均的系数时，自然选一个过零点，选另一个过右端点，因为这两点都是最优点。於是我们得到两条新曲线 $\underline{\eta}$ 和 $\bar{\eta}$ ，依赖于 α 和 d ，满足 $\underline{\eta} \leq \lambda_0 \leq \bar{\eta}$ 。它们都是 $\bar{k}^{-1}$ 和 $\underline{k}^{-1}$ 的凸平均：

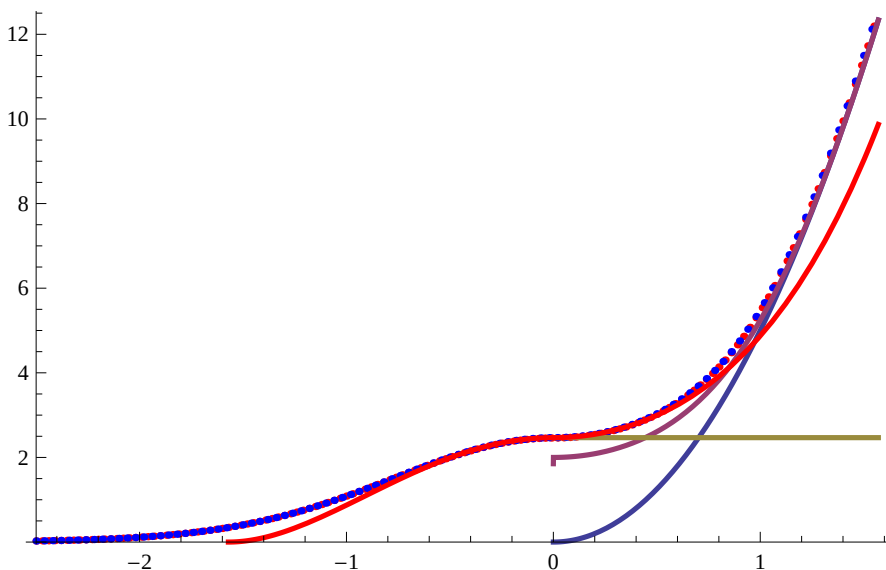
$$\eta = \gamma \bar{k}^{-1} + (1 - \gamma) \underline{k}^{-1}$$

$$\begin{aligned} \bar{\gamma} &= \frac{5^{3/2} - 5\pi^2/16}{5^{3/2} - 10/3} \approx 0.39 && \boxed{\text{曲线经过 } |\alpha| = 0} \quad (\text{与 } \alpha, d \text{ 无关}) \\ \underline{\gamma} &= \left. \frac{d\pi^2/4 - \underline{k}^{-1}}{\bar{k}^{-1} - \underline{k}^{-1}} \right|_{|\alpha|=\pi/2} && \boxed{\text{曲线经过 } |\alpha| = \frac{\pi}{2}} \quad (\text{只与 } d \text{ 有关}) \\ &\approx 0.367 \quad \text{if } d = 5 \end{aligned}$$

这两个凸平均所画出的图如下：



这两个上、下界，在图形上几乎看不出差别，相差非常小。如果查阅本文的电子版，你会发现这里实际上有红和蓝两条曲线。换句话说，我们差不多把马上要讲到的 $\bar{\lambda} = \lambda_0$ 找出来了。使用三个几何量 d , D 和 K 的估计，差不多也就这样了。下图粗略地展示了我们所取得的进步。经过右端点下方的两曲线分别是著名的 Lichnerowicz (1958) 最优估计及已知的最好结果 Chen-Wang (1997)。中间右方的直线是著名的 Zhong-Yang (1984) 最优估计，而经过中点、位于下方、跨越正负两边的曲线是新近的最佳结果 Shi-Zhang (2007)。右端点和零点是仅有的两点，在那里第一特征值有显式精确解。然而，我们的新曲线在所有点上，都近乎精确。



取得这样的结果是预先难以想像的，回想五、六十年来那么多的努力，可见来之不易。在这个过程中，只有最后一步做凸平均时用到特定的几何行为，其它结论都适用于远为一般的情形。今天所讲的所有内容都是一维的，至于怎么样从高维化为一维，下面的比较定理完全是用概率方法做出来，是非常概率的 (参见附录)。

定理 3.4 (比较定理). 我们有

$$\lambda_1 \geq \frac{4}{D^2} \bar{\lambda} \quad (\text{陈和王凤雨 1994: 耦合方法}),$$

此处 $\bar{\lambda}$ 是算子 $\bar{L}$ 的主特征值:

$$\bar{L} = \frac{d^2}{dx^2} + (d-1)\alpha \tanh(\alpha x) \frac{d}{dx}, \quad f(0) = 0, \quad f'(1) = 0.$$

其次，我们有

$$\bar{\lambda} = \lambda_0 \quad (\text{陈 2011: 对偶方法}),$$

其中 λ_0 是算子 L 的主特征值:

$$L = \frac{d^2}{dx^2} - (d-1)\alpha \tanh(\alpha x) \frac{d}{dx}, \quad f'(0) = 0, \quad f(1) = 0.$$

由以上应用也许可以看出，概率这门学科现在已经充分地长大，如果一个数学系里面没有概率，是非常危险的。这话我不只在这里说，早年对大陆的名校的朋友也说过。以上的所有这些方法，对几何都有用。我

讲的是紧流形, 没有边界, 就算是凸边界 (convex boundary) 也是一样。事实上, 就算是非紧流形, 也类似, 只是需要添加一个向量场。我看到最新的是一个 Finsler-Laplacian 估计, 结构也是一样, 只不过要用不同记号而已。谢谢大家。

附录: 耦合方法与概率度量

这次演讲集中于这两年的新进展, 几乎未涉及概率方法, 后者以前访问台湾时曾多次讲过. 这里略作介绍, 以飨读者.

A.1 测度的耦合. 回顾在学微积分学的重积分的时候, 总是利用累次积分, 化成比较容易处理的单重积分来计算. 但也有特别的例外, 把单重积分提升为多重积分来计算. 最典型的例子是计算正态分布

$$\int_{\mathbb{R}} e^{-x^2/2} dx = \sqrt{2\pi},$$

为计算左方的单重积分. 我们先将它提升为二重积分

$$\int_{\mathbb{R}} \int_{\mathbb{R}} e^{-(x^2+y^2)/2} dx dy,$$

此时使用极坐标很容易算出结果. 这个巧妙的算法属于 Simón-Denis Poisson (1781–1840). 换言之, 我们把单个测度 dx 提升为乘积测度 $dx \times dy$. 也可以说, 我们把单个概率测度 (正态分布) $\mu(dx)$ 提升为乘积空间 $\mathbb{R} \times \mathbb{R}$ 上的乘积概率测度 $\mu(dx) \times \mu(dy)$. 这已经是一种特殊的耦合, 称为独立耦合.

想想为什么这种耦合的应用很有限, 原因就在于“独立性”的要求太强了. 只要脱离独立性, 容许相关性, 那就是一片广阔新天地. 给定两个可测空间 $(E_k, \mathcal{E}_k)$ 上的概率测度 $\mu_k, k = 1, 2$, 称乘积空间 $(E_1 \times E_2, \mathcal{E}_1 \times \mathcal{E}_2)$ 上的概率测度 $\tilde{\mu}$ 为 μ_1 和 μ_2 的耦合, 如果下述边缘性条件成立:

$$\begin{aligned} \tilde{\mu}(A_1 \times E_2) &= \mu_1(A_1), & A_1 \in \mathcal{E}_1, \\ \tilde{\mu}(E_1 \times A_2) &= \mu_2(A_2), & A_2 \in \mathcal{E}_2. \end{aligned} \tag{M}$$

近乎平凡的耦合是独立耦合: $\tilde{\mu}_0 = \mu_1 \times \mu_2$. 它却有非平凡的应用. 我们断言直线上的每一个概率测度 μ , 总满足如下的 **FKG** 不等式:

$$\int_{\mathbb{R}} f g d\mu \geq \int_{\mathbb{R}} f d\mu \int_{\mathbb{R}} g d\mu, \quad f, g \in \mathcal{M},$$

其中 $\mathcal{M}$ 是 $\mathbb{R}$ 上有界单调增函数的全体. 使用独立耦合 $\tilde{\mu}_0 = \mu \times \mu$ 和直线的全序性, 证明只需一行:

$$\iint_{\mathbb{R} \times \mathbb{R}} \tilde{\mu}_0(dx, dy)[f(x) - f(y)][g(x) - g(y)] \geq 0, \quad f, g \in \mathcal{M}.$$

展开左方的积分, 即得所求.

下面是一种很有用的基本耦合 $\tilde{\mu}_b$. 设 $E_k = E$ (Hausdorff 空间), $k = 1, 2$. 记 Δ 为 E 的对角线集: $\Delta = \{(x, x) : x \in E\}$. 命

$$\tilde{\mu}_b(dx_1, dx_2) = (\mu_1 \wedge \mu_2)(dx_1)I_\Delta + \frac{(\mu_1 - \mu_2)^+(dx_1)(\mu_1 - \mu_2)^-(dx_2)}{(\mu_1 - \mu_2)^+(E)}I_{\Delta^c},$$

其中 $\nu^\pm$ 为符号测度 ν 的 Jordan-Hahn 分解, 而 $\nu_1 \wedge \nu_2 = \nu_1 - (\nu_1 - \nu_2)^+$. 特别地, 如 μ_k 关于 Lebesgue 测度有密度 h_k , 则 $(\mu_1 - \mu_2)^\pm$ 和 $\mu_1 \wedge \mu_2$ 亦然: 分别有密度 $(h_1 - h_2)^\pm$ (函数的正、负部) 和 $h_1 \wedge h_2 := \min\{h_1, h_2\}$. 每一种耦合都有它的基本特征. 这里是上式中的第一项. 有了它之后, 由边缘性自然导出上式中的第二项. 因为 $(\mu_1 - \mu_2)^+$ 表示 μ_1 比 μ_2 多出的部分, $\mu_1 \wedge \mu_2$ 是最大的共有部分. 此耦合的第一项就是把这个共有部分都放到对角线上. 特别地, 如果 $\mu_1 = \mu_2$, 则第二部分消失. 基本耦合的重要性在于它刻画了概率测度的全变差距离. 详言之, 暂设 ρ 为离散距离:

$$\rho(x, y) = \begin{cases} 1 & x \neq y, \\ 0 & x = y. \end{cases}$$

则我们有如下的 Dobrushin 定理:

$$\int_{E \times E} \rho d\tilde{\mu}_b =: \tilde{\mu}_b(\rho) = \inf_{\tilde{\mu} \in \mathcal{C}(\mu_1, \mu_2)} \tilde{\mu}(\rho) = \frac{1}{2} \|\mu_1 - \mu_2\|_{\text{Var}} \quad (= (\mu_1 - \mu_2)^+),$$

其中 $\mathcal{C}(\mu_1, \mu_2) := \{\tilde{\mu} : \tilde{\mu} \text{ 是 } \mu_1 \text{ 和 } \mu_2 \text{ 的耦合}\}$. 对于一般的距离函数 ρ , 如下的定义

$$W(\mu_1, \mu_2) = \inf_{\tilde{\mu} \in \mathcal{C}(\mu_1, \mu_2)} \tilde{\mu}(\rho)$$

称为 μ_1 和 μ_2 的 **Wasserstein** 距离. 这是 1969 年在研究随机场 (概率论与统计物理的交叉学科) 时引进的, 随后逐步成为研究诸多数学课题 (特别是无穷维) 的基本工具.

进一步, 将上述距离函数 ρ 换成更一般的费用函数 c (非负):

$$\inf_{\tilde{\mu} \in \mathcal{C}(\mu_1, \mu_2)} \int_{E \times E} c(x, y) \tilde{\mu}(dx, dy),$$

并考察输运函数 $T: E \rightarrow E$, 使得 μ_2 是在 μ_1 之下、由 T 导出的分布, 而且 $\int_E c(x, T(x)) \mu_1(dx)$ 达到上式的下确界, 这就构成了当今 PDE 中很热闹的研究方向: 最优输运 (optimal transport).

至此为止, 我们讲的是静态情形 (即不含时间 t). 下面转入动态情形. 对于给定的两个马氏过程, 依照上述方法, 我们可以定义它们对于固定的不同出发点、在固定的同一时间 t 的分布 (概率测度) 的耦合. 然而, 这种耦合未必是马氏的. 虽然非马氏耦合有它的用处, 但为简单计, 此处我们只考虑马氏耦合 (即给定边缘过程是马氏的, 要求耦合过程也是马氏的).

A.2 马氏过程的耦合. 给定 $(E_k, \mathcal{E}_k)$ 的马氏半群 $P_k(t)$ 或转移概率 $P_k(t, x_k, \cdot)$, $k = 1, 2$, 称乘积空间上的马氏半群 $\tilde{P}(t)$ 或转移概率 $\tilde{P}(t; x_1, x_2; \cdot)$ 是 $P_k(t)$ 或 $P_k(t, x_k, \cdot)$ ($k = 1, 2$) 的耦合, 如下述 (关于过程的) 边缘性成立:

$$\begin{aligned}\tilde{P}(t; x_1, x_2; A_1 \times E_2) &= P_1(t, x_1, A_1), \\ \tilde{P}(t; x_1, x_2; E_1 \times A_2) &= P_2(t, x_2, A_2), \quad t \geq 0, x_k \in E_k, A_k \in \mathcal{E}_k, k = 1, 2.\end{aligned}\tag{MP}$$

等价地,

$$\begin{aligned}\tilde{P}(t)f(x_1, x_2) &= P_1(t)f(x_1), \quad f \in {}_b\mathcal{E}_1 \\ \tilde{P}(t)f(x_1, x_2) &= P_2(t)f(x_2), \quad f \in {}_b\mathcal{E}_2, t \geq 0, x_k \in E_k, k = 1, 2,\end{aligned}\tag{MP}$$

此处 ${}_b\mathcal{E}$ 是有界 $\mathcal{E}$ 可测函数的全体; 在等式的左边, 原为单变量的函数 f 都被视为双变量函数.

上述定义 “(MP)” 实际上没有多少用处, 因为 $P_k(t)$ 或 $P_k(t, x_k, \cdot)$ ($k = 1, 2$) 都是未知的. 然而, 从它们出发, 自然导出如下关于算子的耦合.

A.3 马氏过程的耦合. 分别以 L_k ($k = 1, 2$) 和 $\tilde{L}$ 表边缘半群 $P_k(t)$ ($k = 1, 2$) 和耦合半群 $\tilde{P}(t)$ 的形式无穷小生成元, 在 (MP) 两边关于 t 在 0 处取导数, 得出如下 (关于算子) 的边缘性:

$$\begin{aligned}\tilde{L}f(x_1, x_2) &= L_1f(x_1), \quad f \in \mathcal{F}_1, \\ \tilde{L}f(x_1, x_2) &= L_2f(x_2), \quad f \in \mathcal{F}_2, x_k \in E_k, k = 1, 2;\end{aligned}\tag{MO}$$

其中 $\mathcal{F}_k$ 是 ${}_b\mathcal{E}_2$ 的适当子集; 与前面的 (MP) 一样, 等式左边的函数 f 都被视为双变量函数.

现在, 对于马氏过程的马氏耦合, 我们有了算子耦合的可行手段. 例如, 对于 $\mathbb{R}^d$ 上的扩散过程, 我们有二阶微分算子

$$L = \sum_{i,j=1}^d a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^d b_i(x) \frac{\partial}{\partial x_i}.$$

为简单计, 写成 $L \sim (a(x), b(x))$. 今给定边缘算子 $L_k \sim (a_k(x), b_k(x))$, $k = 1, 2$, 则由耦合算子的边缘性, 乘积空间 $\mathbb{R}^d \times \mathbb{R}^d$ 上的耦合算子 $\tilde{L} \sim (a(x, y), b(x, y))$ 应有如下形式:

$$a(x, y) = \begin{pmatrix} a_1(x) & c(x, y) \\ c(x, y)^* & a_2(y) \end{pmatrix}, \quad b(x, y) = \begin{pmatrix} b_1(x) \\ b_2(y) \end{pmatrix},$$

其中矩阵 $c(x, y)^*$ 是 $c(x, y)$ 的转置, 条件是保证 $a(x, y)$ 非负定. 更具体些, 设 $L_1 = L_2 \sim (a(x), b(x))$, $a(x) = \sigma(x)\sigma(x)^*$, $\det \sigma \neq 0$. 则可取

$$c(x, y) = \sigma(x)[I - 2\bar{u}\bar{u}^*]\sigma(y)^*, \quad x \neq y,$$

其中 $\bar{u} = (x - y)/|x - y|$. 矩阵 $I - 2\bar{u}\bar{u}^*$ 是反射矩阵 (即行列式为 -1 的正交矩阵). 这个耦合称为反射耦合.

A.4 应用于特征值估计. 回顾 $\{P_t\}_{t \geq 0}$ 与自身的耦合 $\{\tilde{P}_t\}_{t \geq 0}$ 满足边缘性: 对于一切 $f \in C_b^2(\mathbb{R}^d)$ 及 (x, y) ($x \neq y$), 有

$$\tilde{P}_t f(x, y) = P_t f(x) \quad (\text{相应地, } \tilde{P}_t f(x, y) = P_t f(y)), \quad (\text{A1})$$

此处还是将 f 视为双变量函数. 我们由此出发来证明特征值估计.

第一步. 设 g 是相应于 λ_1 的算子 $-L$ 的特征函数: $-Lg = \lambda_1 g$. 则由关于半群的标准微分方程 (即 Kolmogorov 向前微分方程), 我们有

$$\frac{d}{dt} P_t g(x) = P_t L g(x) = -\lambda_1 P_t g(x).$$

固定 g 和 x , 解这个关于函数 $t \rightarrow P_t g(x)$ 的常微分方程, 得出

$$P_t g(x) = g(x) e^{-\lambda_1 t}, \quad t \geq 0. \quad (\text{A2})$$

这个恒等式很漂亮, 它将特征值、特征函数和半群三者统一到一个简洁的公式里. 可惜此刻这个公式无用, 因为三者均未知.

第二步. 幸运的是: 耦合方法使得上述公式变得强有力. 我们所需要的只是关于耦合算子的下述估计:

$$\tilde{L}\rho(x, y) \leq -\alpha\rho(x, y), \quad x \neq y, \quad (\text{A3})$$

其中 $\alpha > 0$ 为常数. 它等价于

$$\tilde{P}_t \rho(x, y) \leq \rho(x, y) e^{-\alpha t}, \quad t \geq 0. \quad (\text{A4})$$

今考虑紧空间情形. 此时 g 关于 ρ Lipschitz 连续. 记其 Lipschitz 常数为 c_ρ . 命 $g_1(x, y) = g(x)$, $g_2(x, y) = g(y)$. 那么, 我们有

$$\begin{aligned} e^{-\lambda_1 t} |g(x) - g(y)| &= |P_t g(x) - P_t g(y)| \quad (\text{由 (A2)}) \\ &= |\tilde{P}_t g_1(x, y) - \tilde{P}_t g_2(x, y)| \quad (\text{由 (A1)}) \\ &= |\tilde{P}_t (g_1 - g_2)(x, y)| \quad (\text{因耦合在同一乘积空间上}) \\ &\leq \tilde{P}_t |g_1 - g_2|(x, y) \\ &\leq c_\rho \tilde{P}_t \rho(x, y) \quad (\text{Lipschitz 性}) \\ &\leq c_\rho \rho(x, y) e^{-\alpha t} \quad (\text{由 (A4)}). \end{aligned}$$

因为 g 非常数, 存在 $x \neq y$ 使得 $g(x) \neq g(y)$. 固定这样一对 (x, y) , 然后令 $t \rightarrow \infty$, 必然得出 $\lambda_1 \geq \alpha$. 经过这么简单的两步, 我们就得到了所需的下界估计.

留心上面的紧性假定是可以避免的, 只需使用局部化程序. 因此, 我们的方法适用于非常一般的情形. 作为示例, 考虑 O.U. 过程:

$$L = \sum_{i=1}^d \left(\frac{\partial^2}{\partial x_i^2} - x_i \frac{\partial}{\partial x_i} \right).$$

取 $\rho(x, y) = |x - y|$ (通常的欧氏距离), 稍作计算, 便可看出条件 (A3) 对于反射耦合及 $\alpha = 1$ 成立, 此时达到精确估计: $\lambda_1 = 1 = \alpha$. 应当指出, 在一般情况下, 由条件 (A3) 可见, 为得到好的下界 α , 不仅需要选择好的耦合, 还需要选取好的距离. 不难想象, 这里有很多故事.

多年前, 我们曾经感到很奇怪, 为什么文献上处理的几乎都是 Dirichlet 边界情形, 后来才明白这种情形等价于极大值原理: 其特征函数在内部为正、在边界上达到最小值 0. 对于 λ_1 , 特征函数的零值曲面在区域内部, 极大值原理不适用. 所以, 这里的关键点是用耦合方法这一概率论工具代替了极大值原理.

致谢 作者感谢中研院数学所的邀请 (hosts: 黄启瑞、许顺吉), 感谢黄、许、姜祖恕、周云雄和他们夫人们的热情款待. 感谢陈隆奇、徐洪坤、罗梦娜、郭美惠、李育嘉及夫人、许元春和胡殿中等教授及其团队的邀请和热情款待. 乘此机会, 也感谢数学所的其他同仁, 秘书组、图书馆等行

政团队所提供的帮助和温暖如家的环境，感谢陈丽伍助理为整理我的几篇文章所付出的辛劳。同时，作者感谢国家自然科学基金重点项目 (No. 11131003) 和教育部 973 项目的资助.

参考文献

- [1] Chen, Mu-Fa (2011). *General estimate of the first eigenvalue on manifolds*. Front. Math. China 6(6): 1025-1043.
- [2] Chen, Mu-Fa (2013). *Bilateral Hardy-type inequalities*. Acta Math. Sin., Eng. Ser. 29, no. 1, 1-32.