蛙鸣

中国科大学生数学杂志

1981.6-2023.7

第66期

喜聽蛙鳴

科大學風淳
蛙鳴顯精神
前輩栽好樹
更有後來人

單博
辛丑八月

院长寄语

《蛙鸣》承载着科大数院的同学们探索数学、分享知识、创造新天地的满腔热情和无限期待。一声声纯粹而质朴的“蛙鸣”，其形虽小，其声也宏。值此《蛙鸣》创刊 40 周年之际，希望同学们以此为新的起点，携手奋进，将科大数学的精神薪火相传，在数学世界里继续追寻纯粹崇高的理想，求实求真，勇于创新，奏响科大数学人的最强音！

马志明

2021 年 6 月 4 日

卷首语

我们挚爱着这片数学的沃土
期盼用自由的探讨、分享、交流与创造
为这大地撒下属于我们的理想之花

我们激荡着对于数学和世界的独特见解
感受着那初心与激情相碰撞的火花
那如同青春般纯粹而质朴的蛙鸣啊
在我们每一个人的内心深处迸发

1981 至 2023，《蛙鸣》在 42 年间见证了昔日的小作者、小读者在学术与科研的沃土中生根发芽。

2021 年 6 月 20 日，一声蛙鸣打破了十年沉寂，正值《蛙鸣》40 周年之际，第 64 期《蛙鸣》与大家见面。

2022 年 6 月 26 日，第 65 期《蛙鸣》如约而至，我们一起欢腾、一起感悟，凝聚成盛夏的回忆。

2023 年盛夏，第 66 期《蛙鸣》再赴约，让我们一起“咕呱咕呱”，一起“哇啦哇啦”，在这个夏日展现属于我们的“蛙鸣”！

第 66 期《蛙鸣》编辑部
2023 年 7 月 4 日

目录

初阳

蛙声一片

笔墨诗篇

皇榜


初
阳
阳

稻花香里说丰年，听取蛙声一片。

——《西江月 · 夜行黄沙道中》

从 $(xI_n - A)$ 的列变换矩阵求 A 的标准型基底

安金鹏　林开亮　孙亦青

摘 要

对有限维向量空间 V 上的线性变换 T, 我们提出了构造 V 的基的新算法, 使得 T 在该基下的矩阵表示为 Jordan 标准型和有理标准型. 这个算法给出的基与一个经典算法给出的基吻合. 然而, 此处的算法看起来更直接有效.

1 引言

设 V 是域 F 上的有限维向量空间, T 是 V 上的线性变换. 众所周知, 求 T 的有理标准型和 Jordan 标准型[1] 的问题, 归结为不变因子的计算. 而不变因子可以通过初等变换计算. 令 A 是 T 在 V 的一组基 $\{v_1, \ldots, v_n\}$ 下的矩阵, 这里 $n = \dim V$. 对多项式矩阵 $xI_n - A$ 做初等行列变换, 可以得到 Smith 标准型. 依据熟知的算法, 可以求出可逆多项式矩阵 $P, Q \in \mathrm{GL}_n(F[x])$ 使得 $xI_n - A$ 化为 Smith 标准型:

$$P(xI_n - A)Q = \mathrm{diag}(1, \ldots, 1, d_1, \ldots, d_r). \tag{1.1}$$

其中 $d_1, \ldots, d_r \in F[x]$ 是不变因子, 满足 $d_1|d_2|\cdots|d_r$. 有了不变因子, 就可以读出 A 的有理标准型. 将各个不变因子完全分解, 就可以读出 A 的 Jordan 标准型.

至此, 尚未给出 V 的一组基, 使得 T 在该基下的矩阵表示为有理标准型或 Jordan 标准型. 为此目的, 需要做更多的工作. 文献中已经给出了一些算法, 例如 [1, 3, 4] 描述了一种经典的算法, 从 $F[x]$-模的观点来看, 这也许是最自然的算法.

本文旨在给出一种新算法来求这些基, 其出发点依然是多项式矩阵 $xI_n - A$ 的 Smith 标准型. 与上述算法不同, 我们主要用到通过行列初等变换将 $xI_n - A$ 对角化的列变换矩阵.

2 Jordan 标准型的新算法

我们首先考虑 Jordan 标准型, 从而假定各个不变因子 $d_1, \ldots, d_r$ 在 F 上完全分裂.

为简单起见, 通过对 V 中向量取基底 $\{v_1, \ldots, v_n\}$ 下的坐标, 将 V 等同于 F^n. 于是 T 等同于 F^n 上由 A 的左乘给出的线性变换 L_A. 我们的构造将用到出现在 (1.1)中的多项式矩阵 Q 的各个列向量. 于是我们记

$$Q = (*, \ldots, *, \xi_1, \ldots, \xi_r), \qquad \xi_i \in F[x]^n. \tag{1.2}$$

注意到 Q 可以通过对增广矩阵 $\begin{pmatrix} xI_n - A \\ I_n \end{pmatrix}$ 做初等行列变换得到, 其中行变换只对矩阵 $xI_n - A$ 的 n 行做. 当 $xI_n - A$ 化成 Smith 标准型时, I_n 就变成 Q.

作者信息: 安金鹏, 北京大学数学科学学院教授. 林开亮, 西北农林科技大学理学院讲师, 邮箱 kailiang_lin@163.com. 孙亦青, 北京大学数学科学学院 2020 级.

稿件信息: 2022 年 12 月 8 日来稿, 2023 年 6 月 15 日修改, 6 月 19 日接受, 稿件编辑: 曾相如.

[1] 但凡涉及 Jordan 标准型, 我们总假定 T 的特征多项式在 F 上完全分裂.

对 A 的每个特征值 $\lambda \in \mathrm{Spec}(A)$, 令它在 $d_1, \ldots, d_r$ 中的重数依次为 $m_1(\lambda), \ldots, m_r(\lambda)$. 于是我们有

$$d_i = \prod_{\lambda \in \mathrm{Spec}(A)} (x - \lambda)^{m_i(\lambda)}, \qquad 1 \leqslant i \leqslant r, \tag{1.3}$$

且

$$\chi_A(x) = \det(xI_n - A) = \prod_{i=1}^{r} d_i = \prod_{\lambda \in \mathrm{Spec}(A)} (x - \lambda)^{m(\lambda)}, \tag{1.4}$$

其中

$$m(\lambda) = m_1(\lambda) + \cdots + m_r(\lambda) \tag{1.5}$$

为特征值 λ 的代数重数.

为给出一组基, 使得它可以给出 Jordan 标准型, 我们先引入一些记号.

- 对多项式 $f \in F[x]$ 以及 $c \in F$ 与整数 $j \geqslant 0$, 令 $\langle f, (x-c)^j \rangle$ 表示 f 中关于 $x - c$ 的展开式中 $(x-c)^j$ 的系数.
- 对一个多项式列向量 $\xi = (f_1, \ldots, f_n)^t \in F[x]^n$ (上标 t 表示转置) 以及 $c \in F$ 与整数 $j \geqslant 0$, 记 $\langle \xi, (x-c)^j \rangle = (\langle f_1, (x-c)^j \rangle, \ldots, \langle f_n, (x-c)^j \rangle)^t$.

我们对 Jordan 基的算法如下.

定理 2.1. 设 $P, Q \in \mathrm{GL}_n(F[x])$ 满足 (1.1), 将 Q 记为 (1.2), 假定给出了分解 (1.3) . 对每个 $\lambda \in \mathrm{Spec}(A)$ 以及每个使得 $m_i(\lambda) \neq 0$ 的 $i \in \{1, \ldots, r\}$, 令

$$\alpha_{ij}(\lambda) = \langle \xi_i, (x-\lambda)^j \rangle \in F^n, \qquad 0 \leqslant j \leqslant m_i(\lambda) - 1. \tag{1.6}$$

则以下断言成立:

(i) 向量

$$\{\alpha_{ij}(\lambda) : 0 \leqslant j \leqslant m_i(\lambda) - 1\} \tag{1.7}$$

线性无关, 并生成 F^n 的一个 L_A-不变子空间 $W_i(\lambda)$. 限制线性变换 $L_A|_{W_i(\lambda)}$ 在 $W_i(\lambda)$ 的基底 (1.7) 下的矩阵对应于特征值 λ 的 m_i 阶 Jordan 块:

$$J_{m_i}(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix}. \tag{1.8}$$

(ii) F^n 是子空间 $W_i(\lambda)$ 的直和. 从而向量

$$\{\alpha_{ij}(\lambda) : \lambda \in \mathrm{Spec}(A), m_i(\lambda) \neq 0, 0 \leqslant j \leqslant m_i(\lambda) - 1\} \tag{1.9}$$

构成 F^n 的一组基, L_A 在这组基下为 Jordan 标准型.

注 . (i) 若 F 的特征为 0, 则向量 $\alpha_{ij}(\lambda)$ 可以如下求出: $\alpha_{ij}(\lambda) = \frac{1}{j!} \frac{\mathrm{d}^j}{\mathrm{d}x^j} \xi_i(x) \Big|_{x=\lambda}$.

(ii) 显然, 若 M 由 (1.9)中的向量拼成的矩阵, 则它是使得 $M^{-1}AM$ 为 Jordan 标准型的过渡矩阵.

定理 2.1 的证明

(i) 的证明. 根据 (1.1) 以及 (1.2), 我们有

$$P(xI_n - A)\xi_i = (0, \dots, d_i, \dots, 0)^t. \tag{1.10}$$

现在 $(x-\lambda)^{m_i(\lambda)} \mid d_i$, 从而 $(x-\lambda)^{m_i(\lambda)} \mid (0, \dots, d_i, \dots, 0)^t$(在整除每一个分量的意义下), 即有 $(x-\lambda)^{m_i(\lambda)} \mid P(xI_n - A)\xi_i$, 进而 $(x-\lambda)^{m_i(\lambda)} \mid P^{-1}P(xI_n - A)\xi_i = (xI_n - A)\xi_i$. 这就意味着 $(xI_n - A)\xi_i$ 关于 $(x-\lambda)$ 的展开式中, 低于 $m_i(\lambda)$ 次的系数都等于 0. 注意, 根据定义, 我们有

$$\xi_i \equiv \sum_{j=0}^{m_i(\lambda)-1} \alpha_{ij}(\lambda)(x-\lambda)^j \pmod{(x-\lambda)^{m_i}}. \tag{1.11}$$

而

$$(xI_n - A) = -(A - \lambda I_n) + I_n(x-\lambda), \tag{1.12}$$

从而不难算出

$$\langle (xI_n - A)\xi_i, (x-\lambda)^j \rangle = \begin{cases} -(A-\lambda I_n)\alpha_{i0}(\lambda), & j = 0 \\ \alpha_{i,j-1}(\lambda) - (A-\lambda I_n)\alpha_{ij}(\lambda), & 1 \leqslant j \leqslant m_i(\lambda) - 1 \end{cases} \tag{1.13}$$

这就推出

$$\begin{cases} (A-\lambda I_n)\alpha_{i0}(\lambda) = 0, \\ (A-\lambda I_n)\alpha_{ij}(\lambda) = \alpha_{i,j-1}(\lambda), & 1 \leqslant j \leqslant m_i(\lambda) - 1 \end{cases} \tag{1.14}$$

为说明 $\alpha_{i0}(\lambda) \neq 0$, 我们注意到

$$\alpha_{i0}(\lambda) = \xi_i(\lambda) \tag{1.15}$$

而 $\xi_i(\lambda)$ 是可逆矩阵 $Q(\lambda)$ 的第 $n-r+i$ 列.

由此易得 (i) 中结论, 特别地, $L_A|_{W_i(\lambda)}$ 在基底 (1.7) 下的矩阵是 $m_i(\lambda)$ 阶 Jordan 块(1.8). [2]

为证明 (ii). 对每个特征值 λ, 令

$$W(\lambda) = \sum_{i\,:\,d_i(\lambda)=0} W_i(\lambda) \tag{1.16}$$

则显然有

$$W(\lambda) \subset \mathrm{Ker}(A-\lambda)^{m(\lambda)} = V(\lambda). \tag{1.17}$$

由根子空间分解定理[3], 我们有

$$V = \bigoplus_{\lambda} V(\lambda) \tag{1.18}$$

从而为证明结论, 我们只要证明 (1.16) 中的和是直和.

为此, 设

$$\sum_{i\,:\,d_i(\lambda)=0} c_i w_i = 0, \tag{1.19}$$

其中 $w_i \in W_i(\lambda), c_i \in F$.

[2] 在许多教材 (例如 [2, p. 236]) 中, Jordan 块不是写成(1.8), 而是写成它的转置. 从我们的处理来看, 似乎这里的形式更自然.

[3] 参见 [2, p. 211] 定理 12. 注意此处我们其实进一步证明了 $\dim V(\lambda) = m(\lambda)$.

设 w_i 被 $(A-\lambda I_n)$ 零化的指数为 q_i, 即

$$(A-\lambda I_n)^{q_i} w_i = 0, \qquad (A-\lambda I_n)^{q_i - 1} w_i \neq 0 \tag{1.20}$$

注意, $q_i > 1$, 并且 $(A-\lambda I_n)^{q_i-1} w_i$ 是 $\alpha_{i0}(\lambda)$ 的一个非零倍数. 不妨设 $q_1 \geq q_2 \cdots \geq q_s > 1 = q_{s+1} = \cdots = q_r$. 将 (1.19) 两边以 $(A-\lambda I_n)^{q_1-1}$ 作用, 我们得到 $\alpha_{i0}(\lambda)$ 之间的一个线性关系. 然而我们注意到 $\alpha_{i0}(\lambda) = \xi_i(\lambda)$ 是可逆矩阵 $Q(\lambda)$ 的各个列, 从而线性无关. 于是我们得到零化指数为 q_1 的各个向量的系数都等于 0. 从而 (1.19) 中这些项扔掉, 得到一个化简的线性关系. 对这个化简的关系用 $(A-\lambda I_n)^{q_2-1}$ 作用, 得到进一步化简. 最终得到所有系数等于 0.

于是 (1.16) 是直和, 从而有

$$\bigoplus_{\lambda\in\mathrm{Spec}(A)} W(\lambda) = \bigoplus_{\lambda\in\mathrm{Spec}(A)} \bigoplus_{i\,:\,d_i(\lambda)=0} W_i(\lambda).$$

由于右边子空间的维数之和显然是 $\sum_{\lambda\in\mathrm{Spec}(A)} m(\lambda) = \dim V$, 所以它就是全空间 V. 至此, **定理 2.1 证毕**.

注 .
- 从证明可以看出, 为求出 A 的 Jordan 标准型, 我们真正需要的, 只是将 $(xI_n - A)$ 对角化, 而并不需要将它化成 Smith 标准型 (参见例 2.2).
- 不必假定 A 的特征多项式在基域 F 上完全分裂, 本算法的一个变形可以求出 A 在 F 上的全部特征值与特征向量. 结论是:A 在 F 上的特征值恰好是各个 d_i 在 F 上的根, 并且对应的特征向量由 ξ_i 在这个根处的取值给出 (见 [6]).
- 上述证明稍加推广, 即可给出任意域上的方阵 A 的广义 Jordan 标准型的基底.

例 2.1. 设

$$A = \begin{pmatrix} 0 & 1 & 1 & -1 \\ 1 & 0 & -1 & 1 \\ 1 & -1 & 0 & 1 \\ -1 & 1 & 1 & 0 \end{pmatrix},$$

我们依照上述算法来求 A 的 Jordan 标准型以及对应的 Jordan 基.

$$\begin{pmatrix} x & -1 & -1 & 1 \\ -1 & x & 1 & -1 \\ -1 & 1 & x & -1 \\ 1 & -1 & -1 & x \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & -1 & -1 & x \\ -1 & x & 1 & -1 \\ -1 & 1 & x & -1 \\ x & -1 & -1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & x-1 & 0 & x-1 \\ -1 & 0 & x-1 & x-1 \\ x & x-1 & x-1 & 1-x^2 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & -x \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x-1 & 0 & x-1 \\ 0 & 0 & x-1 & x-1 \\ 0 & x-1 & x-1 & 1-x^2 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & -x \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x-1 & 0 & 0 \\ 0 & 0 & x-1 & x-1 \\ 0 & x-1 & x-1 & 2-x-x^2 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & -x-1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x-1 & 0 & 0 \\ 0 & 0 & x-1 & x-1 \\ 0 & 0 & x-1 & 2-x-x^2 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & -x-1 \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x-1 & 0 & 0 \\ 0 & 0 & x-1 & 0 \\ 0 & 0 & x-1 & 3-2x-x^2 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & 1 & -x-2 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x-1 & 0 & 0 \\ 0 & 0 & x-1 & 0 \\ 0 & 0 & 0 & (x-1)(x+3) \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & 1 & -x-2 \end{pmatrix}.$$

从而 A 的不变因子为

$$d_1(x) = d_2(x) = x-1, \quad d_3(x) = (x-1)(x+3).$$

$$Q = (*, \xi_1, \xi_2, \xi_3) = \begin{pmatrix} * & 0 & 0 & 1 \\ * & 1 & 0 & -1 \\ * & 0 & 1 & -1 \\ * & 1 & 1 & -x-2 \end{pmatrix}.$$

A 的谱为 $\mathrm{Spec}(A) = \{1, -3\}$.

对应于特征值 $\lambda = 1$, 我们有

$$\alpha_{10}(1) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \quad \alpha_{20}(1) = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \quad \alpha_{30}(1) = \begin{pmatrix} 1 \\ -1 \\ -1 \\ -x-2 \end{pmatrix}_{x=1} = \begin{pmatrix} 1 \\ -1 \\ -1 \\ -3 \end{pmatrix}.$$

对应于特征值 $\lambda = -3$, 我们有

$$\alpha_{30}(-3) = \begin{pmatrix} 1 \\ -1 \\ -1 \\ -x-2 \end{pmatrix}_{x=-3} = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}.$$

从而令

$$M = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & -1 & -1 \\ 0 & 1 & -1 & -1 \\ 1 & 1 & -3 & 1 \end{pmatrix},$$

就有

$$M^{-1}AM=\mathrm{diag}(1,1,1,-3).$$

例 2.2 (对照 [2] pp. 310–312 例 2)**.** 设

$$A=\begin{pmatrix}0&-1&2\\3&8&-14\\3&6&-10\end{pmatrix}.$$

求 M, 使得 $M^{-1}AM$ 为 Jordan 标准形.

$$\begin{pmatrix}x&1&-2\\-3&x-8&14\\-3&-6&x+10\\1&0&0\\0&1&0\\0&0&1\end{pmatrix}\to\begin{pmatrix}1&x&-2\\x-8&-3&14\\-6&-3&x+10\\0&1&0\\1&0&0\\0&0&1\end{pmatrix}\to\begin{pmatrix}1&0&0\\x-8&-x^2+8x-3&2x-2\\-6&6x-3&x-2\\0&1&0\\1&-x&2\\0&0&1\end{pmatrix}$$

$$\to\begin{pmatrix}1&0&0\\0&-x^2+8x-3&2x-2\\0&6x-3&x-2\\0&1&0\\1&-x&2\\0&0&1\end{pmatrix}\to\begin{pmatrix}1&0&0\\0&-x^2+2x&x\\0&6x-3&x-2\\0&1&0\\1&-x&2\\0&0&1\end{pmatrix}\to\begin{pmatrix}1&0&0\\0&x&-x^2+2x\\0&x-2&6x-3\\0&0&1\\1&2&-x\\0&1&0\end{pmatrix}$$

$$\to\begin{pmatrix}1&0&0\\0&2&-x^2-4x+3\\0&x-2&6x-3\\0&0&1\\1&2&-x\\0&1&0\end{pmatrix}\to\begin{pmatrix}1&0&0\\0&2&-x^2-4x+3\\0&0&x(x+1)^2\\0&0&1\\1&2&-x\\0&1&0\end{pmatrix}\to\begin{pmatrix}1&0&0\\0&2&0\\0&0&\boxed{2x(x+1)^2}\\0&0&2\\1&2&2x^2+6x-6\\0&1&x^2+4x-3\end{pmatrix}$$

于是 A 仅有唯一的不变因子 $d(x)=x(x+1)^2$(在此情形我们略去代表不变因子的下标 i), 显然 A 的特征值为 $0,-1$.

对应于 $\lambda=0$ 有

$$\alpha_0(0)=\begin{pmatrix}2\\2x^2+6x-6\\x^2+4x-3\end{pmatrix}_{x=0}=\begin{pmatrix}2\\-6\\-3\end{pmatrix}.$$

对应于 $\lambda=-1$ 有

$$\alpha_0(-1)=\begin{pmatrix}2\\2x^2+6x-6\\x^2+4x-3\end{pmatrix}_{x=-1}=\begin{pmatrix}2\\-10\\-6\end{pmatrix},\quad\alpha_1(-1)=\left(\frac{\mathrm{d}}{\mathrm{d}x}\begin{pmatrix}2\\2x^2+6x-6\\x^2+4x-3\end{pmatrix}\right)_{x=-1}=\begin{pmatrix}0\\4x+6\\2x+4\end{pmatrix}_{x=-1}=\begin{pmatrix}0\\2\\2\end{pmatrix}.$$

从而, 令

$$M = \begin{pmatrix} \alpha_0(0) & \alpha_0(-1) & \alpha_1(-1) \end{pmatrix} = \begin{pmatrix} 2 & 2 & 0 \\ -6 & -10 & 2 \\ -3 & -6 & 2 \end{pmatrix}$$

就有

$$M^{-1}AM = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}.$$

3 有理标准型的新算法

我们给出求有理标准型的一个平行算法如下 (平行的证明此处从略, 一个间接的证明见下一节).

定理 3.1. 设 $P, Q \in \mathrm{GL}_n(F[x])$ 满足 (1.1), 记 Q 如 (1.2). 对 $1 \leqslant i \leqslant r$, 令 $\deg d_i = n_i$, 递归定义以下多项式向量

$$\xi_{i0} = \xi_i, \qquad \xi_{ij} = x\xi_{i,j-1} - d_i\langle \xi_{i,j-1}, x^{n_i-1}\rangle, \quad 1 \leqslant j \leqslant n_i - 1. \tag{1.21}$$

并令

$$\beta_{ij} = \langle \xi_{ij}, x^{n_i-1}\rangle, \qquad 0 \leqslant j \leqslant n_i - 1. \tag{1.22}$$

则下述断言成立:

(i) 向量

$$\{\beta_{ij}, \quad 0 \leqslant j \leqslant n_i - 1\} \tag{1.23}$$

线性无关, 并生成 F^n 的一个 L_A-不变子空间 W_i. 限制线性变换 $L_A|_{W_i}$ 在 W_i 的基底 (1.23) 下的矩阵是 d_i 的友阵.

(ii) F^n 是子空间 W_i 的直和. 从而向量

$$\{\beta_{ij} : 1 \leqslant i \leqslant r, 0 \leqslant j \leqslant n_i - 1\} \tag{1.24}$$

构成 F^n 的一组基, L_A 在这组基下为有理标准型.

注 . 如同 Jordan 标准型的情况, 将 (1.24) 中向量拼接构成的矩阵 N 是使得 $N^{-1}AN$ 为有理标准型的过渡矩阵.

例 3.1. 设

$$A = \begin{pmatrix} 0 & -1 & 2 \\ 3 & 8 & -14 \\ 3 & 6 & -10 \end{pmatrix}$$

求 N, 使得 $N^{-1}AN$ 为有理标准形. 如前, 我们得到 $d = x(x+1)^2 = x^3 + 2x^2 + x$,

$$\xi = \begin{pmatrix} 1 \\ x^2 + 3x - 3 \\ \frac{1}{2}x^2 + 2x - \frac{3}{2} \end{pmatrix}.$$

于是我们有

$$\xi_0 = \xi = \begin{pmatrix} 1 \\ x^2+3x-3 \\ \frac{1}{2}x^2+2x-\frac{3}{2} \end{pmatrix}, \quad \beta_0 = \begin{pmatrix} 0 \\ 1 \\ \frac{1}{2} \end{pmatrix},$$

$$\xi_1 = x\xi_0 - d\beta_0 = x\begin{pmatrix} 1 \\ x^2+3x-3 \\ \frac{1}{2}x^2+2x-\frac{3}{2} \end{pmatrix} - (x^3+2x^2+x)\begin{pmatrix} 0 \\ 1 \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} x \\ x^2-4x \\ x^2-2x \end{pmatrix}, \quad \beta_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix},$$

$$\xi_2 = x\xi_1 - d\beta_1 = x\begin{pmatrix} x \\ x^2-4x \\ x^2-2x \end{pmatrix} - (x^3+2x^2+x)\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} x^2 \\ -6x^2-x \\ -4x^2-2x \end{pmatrix}, \quad \beta_2 = \begin{pmatrix} 1 \\ -6 \\ -4 \end{pmatrix}.$$

于是, 令

$$N = \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & -6 \\ \frac{1}{2} & 1 & -4 \end{pmatrix},$$

就有

$$N^{-1}AN = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -2 \end{pmatrix}.$$

4 与经典算法的比较

我们先回顾一下 [1, 3, 4] 中给出的经典算法.

不同于本文所给出的算法, 经典算法是利用出现在 (1.1)中的矩阵 P 的逆. 令

$$P^{-1} = (*, \ldots, *, \eta_1, \ldots, \eta_r), \qquad \eta_i \in F[x]^n.$$

我们将 F^n 视为 $F[x]$-模, $f \in F[x]$ 与 $\alpha \in F^n$ 的乘法由 $f\alpha = f(A)\alpha$ 给出. 考虑自由 $F[x]$-模 $F[x]^n$ 与唯一的 $F[x]$-模同态 $\phi : F[x]^n \to F^n$, 它在 F^n 上的限制是恒同. 则

$$\mathrm{Ker}(\phi) = \{(xI_n - A)\zeta : \zeta \in F[x]^n\}.$$

经典算法如下:

- 我们有循环分解 $F^n = \bigoplus_{i=1}^r F[x]\phi(\eta_i)$, 且 $\phi(\eta_i)$ 的零化子是 d_i. 于是, 对每个 $1 \leqslant i \leqslant r$, 向量

$$\{x^j\phi(\eta_i) : 0 \leqslant j \leqslant n_i - 1\} \tag{1.25}$$

  构成 $F[x]\phi(\eta_i)$ 的一组基, 并且 L_A 的限制在这组基下的矩阵为 d_i 的友阵.
- 对每个 $1 \leqslant i \leqslant r$, 给定分解 (1.3), 有本原分解 $F[x]\phi(\eta_i) = \bigoplus_{\lambda : d_i(\lambda)=0} F[x]\beta_i(\lambda)$, 其中 $\beta_i(\lambda) = \frac{d_i}{(x-\lambda)^{m_i(\lambda)}}\phi(\eta_i)$ 有零化子 $(x-\lambda)^{m_i(\lambda)}$. 于是向量

$$\left\{\frac{d_i}{(x-\lambda)^{j+1}}\phi(\eta_i) : 0 \leqslant j \leqslant m_i(\lambda) - 1\right\} \tag{1.26}$$

  构成 $F[x]\beta_i(\lambda)$ 的一组基, 并且 L_A 限制在这组基下的矩阵是一个 $m_i(\lambda)$ 阶 Jordan 块, 对应于特征值 λ.

此算法中, 耗时较多的步骤在于向量 $\phi(\eta_i)$ 的计算, 而它是基于 P^{-1} 的计算. 在 [1, 3] 中 P^{-1} 的计算是通过对 I_n 同步记录施加在 $(xI_n - A)$ 上的行变换的逆而得到. 为从 η_i 计算 $\phi(\eta_i)$, 还需要计算多项式分量 η_i 在矩阵 A 的值.

可以证明, 此处 (1.9) 和 (1.24) 所给出的基底与经典算法给出的基底一致. 我们简要说明如下.

我们首先说明, 对 $1 \leqslant i \leqslant r$ 与 d_i 的根 λ, (1.7)中的向量与 (1.26)中的向量一致, 即

$$\alpha_{ij}(\lambda) = \frac{d_i}{(x-\lambda)^{j+1}}\phi(\eta_i), \qquad 0 \leqslant j \leqslant m_i(\lambda) - 1. \tag{1.27}$$

对固定的指标 i, j , 令

$$\xi_i(\lambda) = (x-\lambda)^{j+1}\zeta_{ij}(\lambda) + \rho_{ij}(\lambda),$$

其中 $\zeta_{ij}(\lambda), \rho_{ij}(\lambda) \in F[x]^n$, 且 $\deg \rho_{ij}(\lambda) \leq j$. 则由 (1.6) 可得 $\langle \rho_{ij}(\lambda), (x-\lambda)^j \rangle = \alpha_{ij}(\lambda)$. 由(1.1), 我们有

$$d_i\eta_i = (xI_n - A)\xi_i = (x-\lambda)^{j+1}(xI_n - A)\zeta_{ij}(\lambda) + (xI_n - A)\rho_{ij}(\lambda). \tag{1.28}$$

由于 $(x-\lambda)^{j+1}$ 整除 d_i, 由 (1.28) 得到, $(x-\lambda)^{j+1}$ 整除 $(xI_n - A)\rho_{ij}(\lambda)$ 的各个分量, 这些分量的次数至多为 $j+1$. 另一方面,

$$\langle (xI_n - A)\rho_{ij}(\lambda), (x-\lambda)^{j+1} \rangle = \langle \rho_{ij}(\lambda), (x-\lambda)^{j+1} \rangle = \alpha_{ij}(\lambda).$$

从而必定有 $(xI_n - A)\rho_{ij}(\lambda) = (x-\lambda)^{j+1}\alpha_{ij}(\lambda)$. 将它代入 (1.28)即得

$$\frac{d_i}{(x-\lambda)^{j+1}}\eta_i = (xI_n - A)\zeta_{ij}(\lambda) + \alpha_{ij}(\lambda).$$

将同态 ϕ 作用在上式两边, 并利用 $(xI_n - A)\zeta_{ij}(\lambda) \in \mathrm{Ker}(\phi)$, 即得 (1.27).

类似地, 我们证明, 对每个 $1 \leqslant i \leqslant r$, (1.23) 中的向量与 (1.25) 中的向量一致, 即

$$\langle \xi_{ij}, x^{n_i - 1} \rangle = x^j\phi(\eta_i), \qquad 0 \leqslant j \leqslant n_i - 1. \tag{1.29}$$

由(1.21)我们有 $d_i \mid \xi_{ij} - x^j\xi_i$, 从而我们可以写 $x^j\xi_i = d_i\zeta_{ij} + \xi_{ij}$, 其中 $\zeta_{ij} \in F[x]^n$. 由 (1.1) 有

$$x^j d_i\eta_i = x^j(xI_n - A)\xi_i = d_i(xI_n - A)\zeta_{ij} + (xI_n - A)\xi_{ij}. \tag{1.30}$$

这就推出 d_i 整除 $(xI_n - A)\xi_{ij}$ 的每个分量, 而这些分量的次数不超过 n_i. 另一方面,

$$\langle (xI_n - A)\xi_{ij}, x^{n_i} \rangle = \langle \xi_{ij}, x^{n_i - 1} \rangle$$

于是 $(xI_n - A)\xi_{ij} = d_i\langle \xi_{ij}, x^{n_i - 1} \rangle$. 将它代入 (1.30), 我们得到

$$x^j\eta_i = (xI_n - A)\zeta_{ij} + \langle \xi_{ij}, x^{n_i - 1} \rangle.$$

将同态 ϕ 作用在上式两边, 并注意到 $(xI_n - A)\zeta_{ij} \in \mathrm{Ker}(\phi)$, 即得 (1.29).

5 结语

最后, 我们分享一段 Knuth[5] 的评论, 希望本文介绍的算法应用于课堂教学:

For three years I taught a sophomore course in abstract algebra, for mathematics majors at Caltech, and the most difficult topic was always the study of "Jordan canonical form" for matrices. The third year I tried a new approach, by looking at the subject algorithmically, and suddenly it became quite clear. The same thing happened with the discussion of finite groups defined by generators and relations; and in another course, with the reduction theory of binary quadratic forms. By presenting the subject in terms of algorithms, the purpose and meaning of the mathematical theorems became transparent.

致谢

本文第二作者感谢上海应用技术大学陈浦胤老师和宁德师范学院蒋剑剑老师与他讨论交流.

参考文献

[1] Adkins, W. , Weintraub, S. (1992). *Algebra: An Approach via Module Theory.* New York: Springer-Verlag.
[2] 北京大学数学系前代数小组, 王萼芳、石生明修订, 《高等代数》, 第五版, 北京: 高等教育出版社. 2019.
[3] Dummit, D. , Foote, R. (2004). *Abstract Algebra,* 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc.
[4] Jacobson, N. (1985). *Basic Algebra I,* 2nd ed. New York: W. H. Freeman and Company.
[5] Knuth, D. E. (1974). Computer science and its relation to mathematics, Amer. Math. Monthly 81: 4, 323–343, DOI: 10. 1080/00029890. 1974. 11993556
[6] 乐小英, 利用矩阵的初等变换对矩阵特征值与特征向量同步求解的再讨论, 《江西电力职工大学学报》, 第 10 卷第 1 期, 1997 年 3 月, 16–18.

一个常微分方程初值问题的最大存在区间分析

刘董

摘 要

本文解决了一个特定 ODE 的最大存在区间问题. 仔细分析了解在尚未解决区域内的结构, 并利用换元法求出的临界曲线确定了解在该区域的存在区间. 该方法有望推广至其他问题.

1 引言与预备知识

1.1 问题引入

我们已经知道, 对于 Cauchy 问题

$$\begin{cases}\dfrac{\mathrm{d}x}{\mathrm{d}t}=f(t,x),\\ x(t_0)=x_0.\end{cases} \tag{1}$$

Peano 定理与 Picard 定理分别给出了当 f 连续与满足 Lipschitz 条件时, 解在局部的存在性与存在唯一性. 此外, 解的延伸定理说明了解在某个"最大存在区间"上的存在性. 具体定理如下:

定理 1.1. (Peano) 设函数 $f(t,x)$ 在矩形区域 $R=[t_0-a,t_0+a]\times[x_0-b,x_0+b]$ 内连续, 则初值问题 (1) 在区间 (t_0-h,t_0+h) 上至少存在一个解. 其中,

$$h=\min\left\{a,\frac{b}{M}\right\},\ M=\sup_{(t,x)\in R}|f(t,x)|.$$

定理 1.2. (Picard) 设函数 $f(t,x)$ 在矩形区域 R 内连续, 且对 y 满足 Lipschitz 条件, 则 (1) 在 (t_0-h,t_0+h) 上存在唯一解. 其中 R,h,a 的定义同定理 1.1.

定理 1.3. (解的延伸定理) 设 D 是平面区域, $(t_0,x_0)\in D$, $f(t,x)$ 在 D 内连续, 且对 x 满足局部 Lipschitz 条件. 则 (1) 的唯一解可延伸至 D 的边界.

我们常常对寻找解的最大存在区间感兴趣. 然而, 对于大多数方程而言, 解的存在区间有限时, 其具体边界常常难以求出. 本文主要研究存在区间的有限性, 即方程的解能否向两侧延伸至正、负无穷. 下面的推论对于研究存在区间的有限性特别有用.

推论 1.4. (解的比较定理) 若对于 D 内的任意初值 (t_0,x_0), (1) 均存在唯一解, $\phi(t)$ 是 D 内的任意一条积分曲线. 则 (1) 的解 $x(t)$ 满足:

$$x(t)\begin{cases}<\phi(t), & \text{当 } x(t_0)<\phi(t_0);\\ >\phi(t), & \text{当 } x(t_0)>\phi(t_0);\\ =\phi(t), & \text{当 } x(t_0)=\phi(t_0).\end{cases} \tag{1.1}$$

定理 1.1, 定理 1.2, 定理 1.3, 推论 1.4的证明见 [1].

作者信息: 刘董, 2021 级数学科学学院, 邮箱: dliu2004@mail.ustc.edu.cn.

稿件信息: 2023 年 2 月 27 日来稿, 6 月 10 日修改, 6 月 18 日接受, 稿件编辑: 王子达.

1.2 具体问题介绍

下面, 我们将分析方程

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = (1-x^2)e^{tx^2}, \\ x(t_0) = x_0. \end{cases} \tag{*}$$

在不同初值下, 解的最大存在区间[1].

2 初步讨论

容易验证, $f(t,x) = (1-x^2)e^{tx^2}$ 在 $\mathbb{R}^2$ 上连续且满足局部 Lipschitz 条件. 因此, 由定理 1.2, (*) 的解存在唯一.

我们还可以利用定理 1.3 和推论 1.4 来研究不同积分曲线的性状. 首先注意到 $x = \pm 1$ 是 (*) 的两个解. 我们作出 (*) 决定的方向场, 如下图 1.1 所示, 并将整个平面划分为 D_1, D_2, D_3, D_4 四部分. 其中, $D_1 = \{(t,x) \mid -1 \le x \le 1\}, D_2 = \{(t,x) \mid x < -1\}, D_3 = \{(t,x) \mid x > -1, t \le 0\}, D_4 = \{(t,x) \mid x > 1, t > 0\}$, 如下图 1.2 所示. 对 4 个区域内的初值 (x_0, t_0) 进行分类讨论, 此前已经得到:

(i) $(x_0, t_0) \in D_1$ 时, $x(t)$ 向两侧延伸至无穷;

(ii) $(x_0, t_0) \in D_2$ 时, $x(t)$ 向左延伸至负无穷, 右行最大存在区间有限;

(iii) $(x_0, t_0) \in D_3$ 时, $x(t)$ 向两侧延伸至无穷;

(iv) $(x_0, t_0) \in D_4$ 时, $x(t)$ 向右延伸至无穷, **左行解的存在区间尚不清楚**.

(i)~(iv) 的结论如图 1.3 所示, 相关证明过程见附录. 我们将要研究的就是 D_4 中初值的左行解.

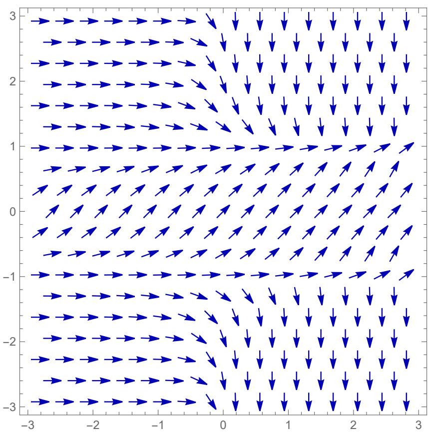


图 1.1: (*) 决定的方向场示意图.

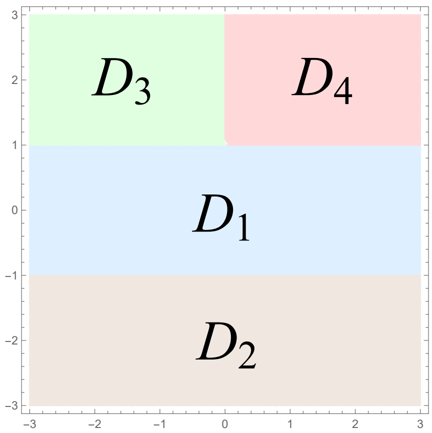


图 1.2: $D_1 \sim D_4$ 所在区域示意图, 具体范围见前文.

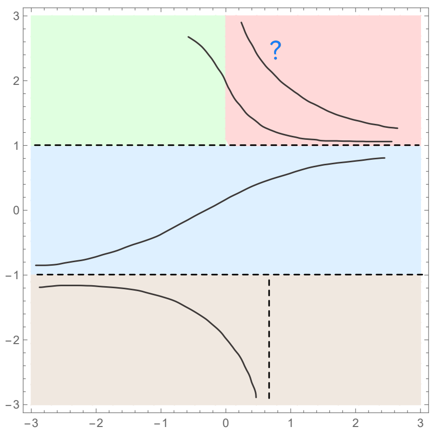


图 1.3: 存在区间示意. 实、虚线分别表示积分曲线和渐近线.

根据上述 (iii), 我们知道当初值位于 D_3 中时, 其积分曲线可以延伸至 D_4. 自然地, 可以得到如下引理:

引理 2.1. 设 $(t_0, x_0) \in D_4$, $x(t)$ 是*的解. 若 $x(t)$ 可延伸至 D_3, 则它可以进一步向左延伸至负无穷.

[1]本题是 2022 秋季学期《微分方程引论》(宁老师班) 附加作业 5 的第 2 题. 本文解决了较为困难的一种情况, 从而完全解决了本题, 并给出了相应的数值解.

证明. 在积分曲线 $x(t)$ 上取一点 $(t_1, x(t_1)) \in D_3$, 并将其作为 (*) 的初值. 由唯一性, 初值问题

$$\begin{cases} \dfrac{\mathrm{d}\tilde{x}}{\mathrm{d}t} = f(t, \tilde{x}), \\ \tilde{x}(t_1) = x(t_1). \end{cases} \tag{1.2}$$

的解 $\tilde{x}(t) = x(t)$. 由 (iii), $\tilde{x}(t)$ 可向左延伸至负无穷, 从而 $x(t)$ 可向左延伸至负无穷. □

3 进一步讨论

3.1 解的结构讨论

引理 2.1 自然引出我们的疑问: 初值在 D_4 中的积分曲线是否都能向左延伸进入 D_3? 如果不能, 根据解的单调性 (D_4 中 $x' = f(t,x) < 0$, 从而单调递减), 这些积分曲线上应当存在奇点. 也就是说, 某些积分曲线以某条直线 $t = A > 0$ 作为垂直渐近线, 并向左、向上延伸至无穷.

考虑 D_4 中全体具有垂直渐近线的积分曲线构成的积分曲线族 Γ(我们先假设 Γ 非空, 后面将直接构造出 Γ 中的元素), 并记 U 为它们扫过的区域, $V = D_4 - U$ 为 D_4 中剩余的部分. 曲线族 Γ 有何结构? 下面的命题回答了这一问题.

命题 3.1. 对于 $\forall \gamma_1, \gamma_2 \in \Gamma$, $\gamma_1 \neq \gamma_2$, $t_0 > 0$, 若 $\gamma_1(t_0) > \gamma_2(t_0)$, 则 $\gamma_1(t) > \gamma_2(t)$ 对任意符合定义的 t 成立.

命题 3.2. 设 $\gamma_1 \in \Gamma$, γ_2 是 (*) 的一条积分曲线, 且 $\exists\, t_0$, 使得 $\gamma_1(t_0) < \gamma_2(t_0)$. 则 $\gamma_2 \in \Gamma$.

证明. 这是引理 2.1 的直接结论. □

换言之, 曲线族 Γ 内的曲线互不相交, 且扫过任意 $\gamma \in \Gamma$ 上方的所有区域. 我们自然得出如下推论:

推论 3.3. 曲线族 Γ 中存在一条临界曲线. 即 $\exists\, \gamma_0 \in \Gamma$, 使得 $\forall \gamma \in \Gamma$, $\gamma \neq \gamma_0$, 有 $\gamma(t) > \gamma_0(t)$ 对任意符合定义的 t 成立.

证明. 任取 $t_0 > 0$. 取 $x_0 = \inf_{\gamma \in \Gamma} \gamma(t_0)$, γ_0 是经过 (t_0, x_0) 的积分曲线. 由命题 3.1, 有 $\forall \gamma \in \Gamma, \gamma(t) > \gamma_0(t)$. 下证 $\gamma_0 \in \Gamma$.

反证. 若 $\gamma_0 \notin \Gamma$, 则 γ_0 可向左延伸至负无穷, 并交 y 轴于点 $(0, \gamma_0(0))$. 令 $\tilde{\gamma_0}$ 为过 $(0, \gamma_0(0) + 1)$ 的积分曲线 (它也可向左延伸至负无穷, 故 $\tilde{\gamma_0} \notin \Gamma$). 但由命题 3.2, $\tilde{\gamma_0}(t_0) > \gamma_0(t_0)$, 从而应有 $\tilde{\gamma_0} \in \Gamma$. 矛盾. □

下面我们的目标就是找出 γ_0. 然而, γ_0 的已知性质过少. 事实上由于没有证明 Γ 非空, 我们甚至还无法确定 γ_0 是否存在！这使得从正面直接求解 γ_0, 即使是求数值解, 相当困难. 作者曾考虑过一种“迂回”思路: 寻找某种迭代所形成的函数列 (类似于 Picard 序列), 使函数列的极限趋于 γ_0. 然而, Picard 序列有其内在的精妙之处, 很难迁移到本问题上来. 作者因此放弃了这种思路.

3.2 换元法

γ_0 难以直接求解的一个重要原因在于, 待求问题并非我们熟悉的初值问题. 然而, 通过简单的换元, γ_0 就可转化为初值问题的解.

考虑变量代换 $u = \frac{1}{x}$, 并将 u 看作自变量. 补全 $u = 0$ 处的定义, 则原方程 (*) 变为

$$\frac{\mathrm{d}t}{\mathrm{d}u} = g(u, t) = \begin{cases} \dfrac{1}{(1 - u^2)} e^{-t/u^2}, & u > 0; \\ \lim\limits_{u \to 0} \dfrac{1}{(1 - u^2)} e^{-t/u^2} = 0, & u = 0. \end{cases} \tag{1.3}$$

我们直接考虑初值问题

$$\begin{cases} \dfrac{\mathrm{d}t}{\mathrm{d}u} = g(u,t), \quad u \ge 0; \\ t(0) = 0 \end{cases} \tag{1.4}$$

的解 $u(t)$(由单调性, $u(t)$ 可由 $t(u)$ 唯一确定). 我们希望证明 $\gamma_0(t) = \frac{1}{u(t)}$, 即 $\frac{1}{u(t)}$ 就是曲线族 Γ 的临界曲线. 首先证明下述引理.

引理 3.4. 初值问题 (1.4) 存在唯一解.

证明. 由于 $g(u,t)$ 在 $(0,0)$ 有一个间断点, 因此无论是 Picard 定理还是 Peano 定理都不能直接应用. 需要另行证明.

存在性: 设 $\alpha > 0$, $t_\alpha(u)$ 是方程 (1.3) 过 $(0,\alpha)$ 的解. $g(u,t)$ 在 $t \ge \alpha > 0, 0 \le u < 1$ 时连续, 且满足局部 Lipschitz 条件. 则由 Picard 定理, $t_\alpha(u)$ 存在且唯一.

下面我们考虑 $\alpha \to 0$ 时函数列 t_α 的极限, 即 $t^*(u) = \lim_{\alpha\to 0} t_\alpha(u)$. $\alpha \to 0$ 时 $t_\alpha(u)$ 恒为正数且单调递减, 故 $t^*(u)$ 逐点收敛. $t^*(u)$ 还符合初值 $(0,0)$. 我们希望证明它就是方程 (1.3) 的一个解. 即证

$$\frac{\mathrm{d}}{\mathrm{d}u}(\lim_{\alpha\to 0} t_\alpha(u)) = g(u, \lim_{\alpha\to 0} t_\alpha(u)), \forall u \in \mathbb{R}. \tag{1.5}$$

该式中, 由连续性

$$\mathrm{RHS} = g(u, \lim_{\alpha\to 0} t_\alpha(u)) = \lim_{\alpha\to 0} g(u, t_\alpha(u)) = \lim_{\alpha\to 0} \frac{\mathrm{d}t_\alpha(u)}{\mathrm{d}u}. \tag{1.6}$$

故只需证

$$\frac{\mathrm{d}}{\mathrm{d}u}(\lim_{\alpha\to 0} t_\alpha(u)) = \lim_{\alpha\to 0} \frac{\mathrm{d}}{\mathrm{d}u}\left(t_\alpha(u)\right). \tag{1.7}$$

这是数学分析中常见的极限号交换问题. t_α 的逐点收敛性已经说明. 从而可知 $g(u,t_\alpha)$ 逐点收敛此外, 容易证明 $\alpha \to 0$ 时 t_α 为单调递减函数列, 于是 $\frac{\mathrm{d}}{\mathrm{d}u}(t_\alpha(u)) = g(u,t_\alpha)$ 单调递增. 从而由 Dini 定理, $\lim_{\alpha\to 0} \frac{\mathrm{d}}{\mathrm{d}u}(t_\alpha(u))$ 内闭一致收敛. 从而极限号可交换, (1.7) 式成立. 存在性证毕.

唯一性: 假设存在 2 个解 t_1, t_2, 且 $t_1(u) \le t_2(u)$. 由方程 (1.4), 有

$$\frac{\mathrm{d}(t_1 - t_2)}{\mathrm{d}u} = \frac{1}{(1-u^2)}\left(e^{-t_1/u^2} - e^{-t_2/u^2}\right) \ge 0. \tag{1.8}$$

注意到 $t_1(0) - t_2(0) = 0$, 从而

$$t_1(u) \ge t_2(u). \tag{1.9}$$

于是 $t_1 = t_2$. 唯一性得证. □

定理 3.5. $\gamma_0(t) = \frac{1}{u(t)}$, 其中 γ_0 是曲线族 Γ 的临界曲线, $u(t)$ 是初值问题 (1.4) 的解.

证明. 直接验证可知 $\frac{1}{u(t)}$ 满足 (*), 且以 $x = 0$ 作为垂直渐近线. 故 $\frac{1}{u} \in \Gamma$. 还需证明它是 Γ 的临界曲线, 即证 $\frac{1}{u(t)}$ 最小.

根据定义, 我们有 $\gamma_0(t) \le \frac{1}{u(t)}, \forall t > 0$. 记 $\tilde{u}(t) = 1/\gamma_0(t)$, 则 $\tilde{u}(t) \ge \frac{1}{u(t)}, \forall t > 0$, 由连续性, $t = 0$ 时该式也成立. 但是, $\gamma_0(t)$ 不能延伸进入第二象限. 形象地说, γ_0 被"夹"在 y 轴和 $\frac{1}{u(t)}$ 之间. 这迫使 $\tilde{u}(0) = 0 = u(0)$. 由引理 3.4 的唯一性, $\tilde{u}(t) = u(t), \forall t \ge 0$, 从而 $\gamma_0(t) = \frac{1}{u(t)}, \forall t > 0$. □

到此为止, 我们已从理论上全部解决了 (*) 的最大存在区间问题. 对于先前未完全解决的情形 (iv), 存在区域 D_4 中的一条直线 γ_0 将 D_4 划分为 2 个部分 U, V. 初值位于 V 中时, 解可向左延伸入 D_3, 并进一步延伸至负无穷; 初值位于 U 中时, 解的左行最大存在区间有限, 积分曲线以某条直线 $t = a > 0$ 作为垂直渐近线.

4 数值求解

下面我们将用数值方法求解出 γ_0 的数值解. 首先利用与上节相同的换元, 求解出初值问题 (1.4) 的解. 然后对作逆变换即得问题的解. 求得的 γ_0 曲线如下文图1.4所示.

具体的 Mathematica 代码见下.

```
u1 = NDSolve[{u'[t] == (1 - u[t]^2)*Exp[t/(u[t]^2)], u[0] == 10^-8}, u, {t, -0.1, 20},
    WorkingPrecision -> 20];
Plot[{Evaluate[1/u[t] /. u1], x = 1}, {t, 0, 0.8}, PlotRange -> {0, 10}, PlotStyle -> {Default, {Gray,
    Dashed}}, AxesLabel -> Automatic]
RegionPlot[{x > Evaluate[1/u[t]] /. u1, t > 0 && x > 1 && x < Evaluate[1/u[t]] /. u1, t < 0 && x > 1,
    -1 < x && x < 1}, {t, -0.2, 0.8}, {x, 0, 7}, BoundaryStyle -> {Thin, Gray}, PlotStyle ->
    {LightRed, Lighter[Lighter[LightRed]], LightGreen, LightBlue}, PerformanceGoal -> "Quality",
    MaxRecursion -> 9(*,WorkingPrecision->20*)]
```

5 总结与延伸

5.1 小结

本文中, 为了完全解决 (*) 的存在区间问题, 我们首先分析了 (*) 的解的结构, 并猜想存在一条临界曲线 γ_0. 在尝试求解临界曲线 γ_0 的过程中, 由于其难以直接求解, 我们利用换元法, 将临界曲线转化为一个初值问题的解, 从而从理论上确定了 γ_0. 随后, 又利用 Mathematica 得到了 γ_0 的数值解.

于是, 对于此前的情况 (iv), 即初值位于 D_4 中的情形, 可以利用临界曲线 γ_0 直接确定左行最大存在区间的有限性, 如图 1.5 所示. 当初值位于 V 中时, 左行最大存在区间有限; 位于 U 中时, 解可向左延伸至无穷.

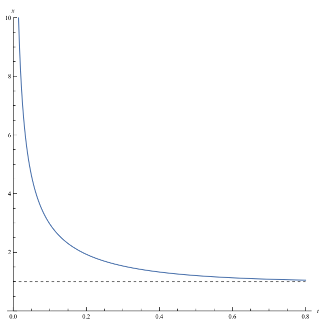


图 1.4: 临界曲线 γ_0 示意图

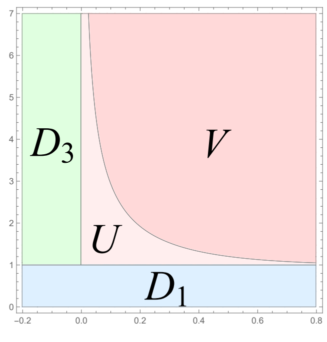


图 1.5: D_4 的划分示意图

5.2 延伸与展望

利用本文的方法, 还有机会判定其他初值问题的解的存在区间.

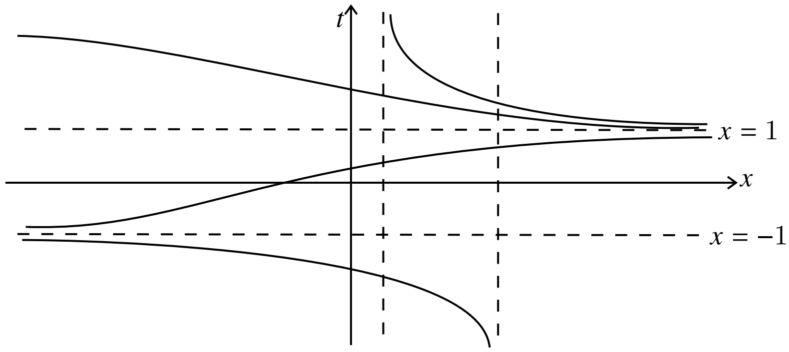
例如, 如果解的存在区间有限, 并且存在奇点. 那么类似地, 可以在所有存在区间有限的解中找出一个临界解, 类似于本文的 γ_0. 本问题中, 一旦求解出 γ_0, 存在区间的问题就解决了. 如果临界解不易直接求解, 可以尝试通过变换 (例如本文的 $u = 1/x$) 将临界解的奇点转换为零点. 这时, 就有机会利用较为简单的初值问题解出临界解, 从而解决存在区间问题.

参考文献

[1] 丁同仁, 李承治. 常微分方程 [M]. 北京: 高等教育出版社, 2004.

A 附录

给出方程 $\frac{dx}{dt} = (1-x^2)e^{tx^2}$ 在情况 (i)-(iv) 的解的最大存在区间.



证明. 显然 $x = \pm 1$ 是方程在 $(-\infty, +\infty)$ 上的解. 由 $|x| > 1$ 时 $(1-x^2)e^{tx^2} < 0$, 即 $\frac{dx}{dt} < 0$ 可得积分曲线在区域 $|x| > 1$ 内严格下降, 类似可得在 $|x| < 1$ 内严格上升.

1. $|x_0| < 1$ 时, 原方程过 (t_0, x_0) 的右行解 $x = \varphi(t)$ 在 $[t_0, +\infty)$ 上满足 $x_0 < x \leqslant 1$, 左行解满足 $-1 \leqslant x < x_0$. 故解的最大存在区间为 $-\infty < t < +\infty$.
2. $x_0 < -1$ 时, 原方程过 (t_0, x_0) 的左行解 $x = \varphi(t)$ 在 $-\infty < t \leqslant t_0$ 上满足 $x_0 \leqslant x \leqslant -1$(由反证易得), 故 $x = \varphi(t)$ 在 $(-\infty, t_0]$ 上存在. 下面证明右行解不能延伸至无穷远. 设右行解为 $x(t)$, 在 $t \geqslant 0$ 且 $t \geqslant t_0$ 时有
$$(1-x^2)e^{tx^2} \leqslant (1-x_0^2)e^{tx^2} \leqslant (1-x_0^2)tx^2.$$
考虑方程 $\frac{d\tilde{x}}{dt} = (1-x_0^2)t\tilde{x}^2$, $\tilde{x}(t_0) = x_0$, 解方程可得
$$\tilde{x}(t) = \frac{2}{(x_0^2-1)(t^2-t_0^2) + \frac{2}{x_0}}.$$
而 $\tilde{x}(t)$ 的解的最大存在区间为 $(-\infty, \sqrt{t_0^2 + 2/(x_0(1-x_0^2))})$, 在 $t \to \sqrt{t_0^2 + 2/(x_0(1-x_0^2))}$ 时 $\tilde{x}(t)$ 趋于负无穷. 由解的比较定理可得 $x(t) \leqslant \tilde{x}(t)$, 故 $x(t)$ 的右行最大存在区间有限.
3. $x_0 > 1$ 时, 类似可得右行解在 $[t_0, +\infty)$ 上存在. 下面考虑左行解的存在区间 J^-.
   (1) $t_0 \leqslant 0$ 时, 由于 J^- 左开右闭, 固定一 $\beta < 0$ 在区间 J^- 中. 任取 $t \leqslant \beta$, 有
   $$|(1-x^2)e^{tx^2}| \leqslant |e^{tx^2}| + \frac{1}{|t|}|tx^2e^{tx^2}| \leqslant 1 - \frac{1}{e\beta}.$$
   (说明: 当 $t \leqslant \beta < 0$ 时, 有 $|e^{tx^2}| \leqslant 1$, $\frac{1}{|t|} < -\frac{1}{\beta}$. 而函数 $g(u) = ue^u$ 在 $(-\infty, 0)$ 上的值域为 $[-\frac{1}{e}, 0)$, 因此 $|tx^2e^{tx^2}| < \frac{1}{e}$.)
   由此可得 $|x(t)| \leqslant |x(\beta)| + (1 - \frac{1}{e\beta})|t-\beta|$, 因此解可延伸至负无穷.
   (2) $t_0 > 0$ 时, 由 (1) 可得此情况下存在 (t_0, x_0) 使得初值问题的解向左延伸至 $-\infty$. 这是因为 (1) 中的积分曲线右行延伸至 $+\infty$, 这条积分曲线上存在 t 使得 $t > 0$ 且 $x(t) > 1$. 以此为初值的积分曲线唯一 (即为上述积分曲线), 可延伸至 $-\infty$. 但下述断言说明: 存在特殊的 $t_0 > 0, x_0 > 1$, 使得左行区间 J^- 有限. □

Borsuk-Ulam 定理的组合应用

叶骁炜

摘 要

本文是笔者大二时阅读 [3] 的读书笔记. 这篇笔记首先介绍一些图论概念, 并给出 Kneser 猜想的叙述 (定理 1.1); 之后, 我们介绍不同版本的 Borsuk-Ulam 定理 (定理 2.1、2.2、2.3 和 2.4), 证明它们的等价性并给出其中一种叙述的一个代数拓扑证明, 所需要的知识可以参考 [1]; 最终, 我们将应用此定理给出 Kneser 猜想的证明.

1 抛出问题

在本文中, $G = (V, E)$ 总是被我们用来表示一个简单无向图, 我们也只考虑简单无向图, 其中 V 是这图的顶点集, 而 E 是它的边集. 图的染色问题往往是很有趣的, 我们这样定义图的**染色**和**染色数**:

定义 1.1. 设 k 是正整数, G 的一个 k-染色是指一个映射 $c : V \to \{1, 2, \cdots, k\}$, 满足

$$\{u, v\} \in E \implies c(u) \neq c(v).$$

G 的染色数, 记为 $\chi(G)$, 是最小的正整数 k, 使得 G 存在 k-染色.

下面我们定义一类特殊的图.

定义 1.2. 设 $n \geq k$ 是两个正整数, 我们称如下的图为 (n, k)-**完全超图**, 记为 $\binom{[n]}{k}$: 其顶点集由 $\{1, 2, \cdots, n\}$ 的所有 k 元子集构成, 两顶点之间有边相连当且仅当这两点对应的 k 元子集不相交.

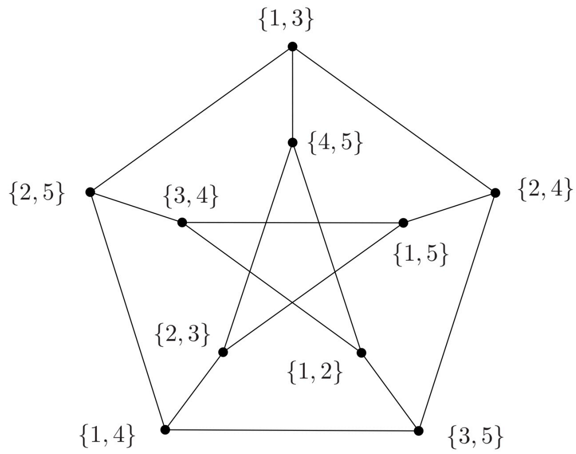
来看几个例子:

例 1.1. (1) 当 $k = 1$ 时, $\binom{[n]}{k}$ 就是 n 阶完全图, 可见完全超图可以视为完全图这一概念的推广. 我们不难验证

$$\chi\left(\binom{[n]}{1}\right) = n.$$

作者信息: 叶骁炜, 2020 级少年班学院, 邮箱: yxwustc@mail.ustc.edu.cn.
稿件信息: 2023 年 3 月 10 日来稿, 3 月 30 日修改, 3 月 30 日接受, 稿件编辑：孙天阳.

(2) $\binom{[5]}{2}$ 具有相当高的知名度, 它的另一个名字叫 Peterson 图.



作为练习, 可以验证

$$\chi\left(\binom{[5]}{2}\right) = 3.$$

我们关心的问题是: 如何计算完全超图的染色数? 具体地说, 能否找到一个函数 f 使得

$$\chi\left(\binom{[n]}{k}\right) = f(n, k)?$$

Kneser 猜想声明:

定理 1.1. 当 $n \geq 2k - 1$ 时, 有

$$\chi\left(\binom{[n]}{k}\right) = n - 2k + 2.$$

这是本文将证明的.

2 Borsuk-Ulam 定理

我们总设 n 是正整数.

定理 2.1 (Borsuk-Ulam, Version I)**.** 任给连续映射 $f : S^n \to \mathbb{R}^n$, 存在 $x \in S^n$, 满足 $f(x) = f(-x)$.

当 $n = 2$ 时, 这个定理告诉我们: 地球上总存在一对对径点, 使得两点处的气温和气压都相同! 这个定理还有许多等价的叙述, 下面列举三个:

定理 2.2 (Borsuk-Ulam, Version II)**.** 不存在连续映射 $f : S^n \to S^{n-1}$, 使得对任意 $x \in S^n$, 均有 $f(x) = -f(-x)$.

定理 2.3 (Borsuk-Ulam, Version III)**.** 设 $S^n = V_1 \cup \cdots \cup V_{n+1}$, 其中每个 V_i 均为闭集, 则它们中必有某个集合包含一组对径点.

定理 2.4 (Borsuk-Ulam, Version IV)**.** 设 $S^n = V_1 \cup \cdots \cup V_{n+1}$, 其中每个 V_i 或为开集, 或为闭集, 则它们中必有某个集合包含一组对径点.

下面我们来验证这几个叙述的等价性.

证明. II $\implies$ I: 用反证法, 假设存在 $f : S^n \to \mathbb{R}^n$ 使得

$$\forall x \in S^n, f(x) \neq f(-x).$$

我们构造一个函数, 使其与定理 2.2 矛盾. 事实上, 令

$$g(x) := \frac{f(x) - f(-x)}{||f(x) - f(-x)||}$$

即可.

I $\implies$ III: 定义连续映射 $f : S^n \to \mathbb{R}^n$ 如下

$$x \in S^n \mapsto (d(x, V_1), \cdots, d(x, V_n)).$$

由定理 2.1 得到, 存在 $x \in S^n$, 满足 $f(x) = f(-x) = y$. 若 y 的第 $i \in \{1, \cdots, n\}$ 个分量为 0, 则 V_i 包含 x 和 $-x$; 否则 V_{n+1} 包含 x 和 $-x$.

III $\implies$ II: 考虑 $\mathbb{R}^n$ 中以原点为中心的 $n-$ 单形的各个面在从原点出发向 S^{n-1} 的投影下的像, 得到: 存在闭集 $F_1, F_2, \cdots, F_{n+1}$ 构成 S^{n-1} 的覆盖, 满足每个 F_i 中均不含对径点.

用反证法, 假设这样的 f 存在, 则 $V_i := f^{-1}(F_i), 1 \leq i \leq n+1$ 便与定理 2.3 所述矛盾.

III $\implies$ IV: 首先来证明 V_i 均是开集的情形. 这时, 每点 $x \in S^n$ 均有一邻域紧包含在某个 V_i 中, 利用 S^n 的紧性, 我们能构造闭集 $F_i \subset V_i$, 使得

$$S^n = \bigcup_{i=1}^{n+1} F_i,$$

由定理 2.3, 存在某个 F_i 包含一对对径点, 那么 V_i 也将包含这对点.

对于一般情形, 我们设 V_i 中有 t 个是闭集, 其余为开集, $0 \leq t \leq n+1$. 我们对 t 归纳证明定理 2.4. $t = 0$ 的情形已经得到证明, 现在我们假设 $t > 0$ 并且闭集个数少于 t 的情形已经得到证明.

此时, 若某个闭集 V_i 包含对径点, 证明也将完成; 若不然, 则这个闭集的直径为 $2 - \varepsilon$, 其中 $\varepsilon > 0$. 我们希望证明其余集合中必有某个包含一组对径点. 为此, 用

$$O := \left\{x \in S^n : d(x, V_i) < \frac{\varepsilon}{2}\right\}$$

代替 V_i, 则由归纳假设, 或者其余集合的某个包含一组对径点, 或者 O 包含对径点. 但三角不等式告诉我们, 后者是不可能的, 证明便完成了.

IV $\implies$ III: 证明平凡. □

注意到 $n = 1$ 时定理 2.1 是连续函数介值定理的直接推论. 下面, 让我们在 $n \geq 2$ 的情况下证明定理 2.2 以结束这一小节的讨论.

证明. 用反证法, 倘若这样的 f 存在, 则它可视为一个 $\mathbb{R}P^n \to \mathbb{R}P^{n-1}$ 的映射. 于是它将诱导映射

$$f^* : H^*(\mathbb{R}P^{n-1}; \mathbb{Z}_2) \to H^*(\mathbb{R}P^n; \mathbb{Z}_2),$$

分析 $H^*(\mathbb{R}P^{n-1};\mathbb{Z}_2)$ 的生成元在 f^* 下的像即知 f^* 是平凡映射. 利用胞腔逼近, 不妨设 f 是胞腔映射, 它在 S^{n-1} 的限制 F(即 $F = i\circ f$, 其中 i 为嵌入) 可视为一个 $\mathbb{R}P^{n-1}\to\mathbb{R}P^{n-1}$ 的映射. 对任意的 k, 设 τ 将 $\mathbb{R}P^k$ 中的胞腔映射到其在 S^k 中的两个原象之和, 那么我们有链复形正合列

$$0\to C_n(\mathbb{R}P^k;\mathbb{Z}_2)\xrightarrow{\tau} C_n(S^k;\mathbb{Z}_2)\xrightarrow{p_\#} C_n(\mathbb{R}P^k;\mathbb{Z}_2)\to 0,$$

其中 $p: S^k\to\mathbb{R}P^k$ 是覆叠映射. 这链复形短正合列给出了同调群长正合列

$$\cdots\to H_n(\mathbb{R}P^k;\mathbb{Z}_2)\xrightarrow{\tau_*} H_n(S^k;\mathbb{Z}_2)\xrightarrow{p_*} H_n(\mathbb{R}P^k;\mathbb{Z}_2)\to H_{n-1}(\mathbb{R}P^k;\mathbb{Z}_2)\to\cdots.$$

易验证由 F 诱导的同态 F_* 与 τ_* 交换, 于是有如下交换图表:

$$\begin{tikzcd}
\cdots \arrow[r] & H_m(\mathbb{R}P^{n-1};\mathbb{Z}_2) \arrow[r, "\tau_*"] \arrow[d, "F_*"] & H_m(S^{n-1};\mathbb{Z}_2) \arrow[r, "p_*"] \arrow[d, "F_*"] & H_m(\mathbb{R}P^{n-1};\mathbb{Z}_2) \arrow[r] \arrow[d, "F_*"] & H_{m-1}(\mathbb{R}P^{n-1};\mathbb{Z}_2) \arrow[r] \arrow[d, "F_*"] & \cdots \\
\cdots \arrow[r] & H_m(\mathbb{R}P^{n-1};\mathbb{Z}_2) \arrow[r, "\tau_*"] & H_m(S^{n-1};\mathbb{Z}_2) \arrow[r, "p_*"] & H_m(\mathbb{R}P^{n-1};\mathbb{Z}_2) \arrow[r] & H_{m-1}(\mathbb{R}P^{n-1};\mathbb{Z}_2) \arrow[r] & \cdots .
\end{tikzcd}$$

注意到 $\mathbb{Z}_2$ 系数下 $\tau_* = 0$ 且 $F_*: H_0(\mathbb{R}P^{n-1};\mathbb{Z}_2)\mapsto H_0(\mathbb{R}P^{n-1};\mathbb{Z}_2)$ 是同构, 由此我们能够证明 $F_*: H_1(\mathbb{R}P^{n-1};\mathbb{Z}_2)\to H_1(\mathbb{R}P^{n-1};\mathbb{Z}_2)$ 也是同构, 于是由泛系数定理自然性知 $F^*: H^1(\mathbb{R}P^{n-1};\mathbb{Z}_2)\to H^1(\mathbb{R}P^n;\mathbb{Z}_2)$ 也是同构, 但 $F^* = i^*\circ f^* = 0$, 矛盾! □

3 问题的解决

我们现在给出定理 1.1 的证明:

证明. 首先证明存在 $(n-2k+2)$-染色, 为此只需将 $F\subset\{1,\cdots,n\}$ 染上编号为 $\min\{F, n-2k+2\}$ 的颜色即可.

我们还需证明不存在 $(n-2k+1)$-染色. 用反证法, 假设存在 $(n-2k+1)$-染色. 记 $d = n-2k+1$.

我们将集合 $\{1,\cdots,n\}$ 的 n 个元素视为 S^d 上一般位置的 n 个点. 这里的“一般位置”是指任一“赤道” S^{d-1} 上点数不超过 d. 对于 $x\in S^d$, 定义 $H(x) := \{y\in S^d : \langle x,y\rangle > 0\}$, 直观地说, 这就是以 x 为中心的半球面. 记

$$A_i := \{x\in S^d : H(x)\ \text{包含第}i\ \text{种颜色的}k\text{-元子集}\},\quad 1\le i\le d,$$

能够证明 A_i 均是开集. 令 $A_{d+1} = (A_1\cup\cdots\cup A_d)^c$, 由定理 2.4, 某个 A_i 中包含一对对径点.

若 $x\in A_i$, $-x\in A_i$, $1\le i\le d$, 则有两个不相交的集合被染成了同一颜色, 这与染色的定义矛盾! 于是, 存在 $x\in S^d$, 满足 $x\in A_{d+1}$, $-x\in A_{d+1}$. 这就表明 $H(x)$ 和 $H(-x)$ 各自至多包含 n 个点中的 $k-1$ 个, 从而赤道 $S^d\setminus(H(x)\cup H(-x))$ 上至少有 $n-2k+2 = d+1$ 个点, 这又与一般位置的假设矛盾! □

注 . 据笔者考据, 公元 2000 年以前的所有 Kneser 猜想的证明均不能绕过 Borsuk-Ulam 定理! 最终, 文献 [4] 给出了首个纯组合证明.

参考文献

[1] Hatcher, A. *Algebraic Topology.* Cambridge University Press, 2002.
[2] Greene, Joshua E. A New Short Proof of Kneser's Conjecture. *The American Mathematical Monthly.* , 109(10), 918-920, Dec. 2002.
[3] Matoušek, J. *Using the Borsuk-Ulam Theorem.* Springer, 2003.
[4] Matoušek, J. A Combinatorial Proof of Kneser's Conjecture. *Combinatorica.* , 24(1), 163-170, 2004.


蛙声一片

稻花香里说丰年，听取蛙声一片。

——《西江月 · 夜行黄沙道中》

Two Compact Hausdorff Spaces are Homeomorphic If and Only If Their Continuous Function Spaces are Isometrically Isomorphic

Zhou Yipeng

Abstract

We prove that the existence of homeomorphism is equivalent to the existence of isometric isomorphism between corresponding continuous function spaces when the topological spaces have some properties. The first part of the article is the proof from Banach about the case on compact metric spaces. His proof uses the directional derivative of the norm to identify peak points of functions, which has not appeared often in the works of other authors. The second part of the article is the proof from Stone about the case on compact Hausdorff spaces.

1 Introduction

We may have find that the topological spaces are closely related to the functions on them. We will introduce works from Banach and Stone. Banach first proved it on compact metric spaces in 1932, and Stone then generalized it to compact Hausdorff spaces in 1937.

2 Body

2.1 On compact metric space

We will first prove Banach-Stone Theorem on compact metric space. Q is a compact metric space, $C(Q)$ means the Banach space of continuous real valued functions defined on Q with the supremum norm.

Lemma 2.1. *Let $f \in C(Q)$ and $s_0 \in Q$. In order that*

$$|f(s_0)| > |f(s)|, \qquad \text{for } \forall s \in Q \text{ with } s \neq s_0, \tag{1.1}$$

it is necessary and sufficient that

$$\lim_{t \to 0} \frac{\|f + tg\| - \|f\|}{t} \tag{1.2}$$

exists for each $g \in C(Q)$.

作者信息: 周奕芃, 2020 级数学科学学院, 邮箱: zhouyp612@mail.ustc.edu.cn.
稿件信息: 2023 年 2 月 26 日来稿, 6 月 10 日修改, 6 月 18 日接受. 稿件编辑: 王子达.

Moreover, if f satisfies (1.1), we have

$$\lim_{t\to 0}\frac{\|f+tg\|-\|f\|}{t}=g(s_0)\operatorname{sgn} f(s_0) \tag{1.3}$$

for each $g\in C(Q)$.

Proof. Let's first show that (1.2) is necessary. If (1.1) holds, then $\|f\|=|f(s_0)|$. Now given $g\in C(Q)$ and a real number t, $f+tg\in C(Q)$, so it attains its maximum absolute value at some $s_t\in Q$. Therefore,

$$|f(s_0)+tg(s_0)|-|f(s_0)|\le\|f+tg\|-\|f\|\le|f(s_t)+tg(s_t)|-|f(s_t)|. \tag{1.4}$$

Also, we have

$$|f(s_0)+tg(s_0)|\le|f(s_t)+tg(s_t)|.$$

So we can get

$$0\le|f(s_0)|-|f(s_t)|\le|t||g(s_0)|+|t||g(s_t)|\le 2|t|\|g\|.$$

It now follows that $\lim_{t\to 0}|f(s_t)|=|f(s_0)|$. And the compactness of Q allow us to conclude that

$$\lim_{t\to 0}s_t=s_0. \tag{1.5}$$

That's because for $\forall n\in\mathbb{N}$, $\exists\delta>0$, such that $\forall t<\delta, |f(s_0)|-|f(s_t)|<\frac{1}{n}$. So we can choose a sequence $\{s_{t_n}\}$ such that $\lim_{n\to\infty}t_n=0$ and therefore $\lim_{n\to\infty}|f(s_{t_n})|=|f(s_0)|$. Since Q is compact, there exists a convergent subsequence $\{s_{t_{n_i}}\}$ and converge to $s\in Q$. Then $|f(s_{t_{n_i}})|$ converge to $|f(s)|=|f(s_0)|$. Because (1.1) holds, we have $s=s_0$. So all the sequences have subsequence converging to s_0, then all the sequences converge to s_0, then $\lim_{t\to 0}s_t=s_0$.

We first suppose that $f(s_0)<0$, by virtue of the fact that $s_t\to s_0$, we may choose t so small that

$$|f(s_0)+tg(s_0)|-|f(s_0)|=-f(s_0)-tg(s_0)+f(s_0)=-tg(s_0)$$

and

$$|f(s_t)+tg(s_t)|-|f(s_t)|=-f(s_t)-tg(s_t)+f(s_t)=-tg(s_t).$$

From the two statement and (1.4) we see that

$$-tg(s_0)\le\|f+tg\|-\|f\|\le -tg(s_t)$$

for sufficiently small t which combined with (1.5) and the continuity of g leads to

$$\lim_{t\to 0}\frac{\|f+tg\|-\|f\|}{t}=-g(s_0).$$

The case where $f(s_0)>0$ (which is the case considered in Banach's book) can be treated in a similiar manner to establish that

$$\lim_{t\to 0}\frac{\|f+tg\|-\|f\|}{t}=g(s_0).$$

This completes the proof of the necessity of the existence of the limit of (1.2) and shows that (1.3) must hold.

For the sufficiency let us assume that $s_0, s_1 \in Q$ with $s_0 \neq s_1$ and

$$\|f\| = |f(s_0)| = |f(s_1)| \geq |f(s)|$$

for all $s \in Q$. If $f(s_0) < 0$, define $g(s) \hat{=} -d(s, s_1)$, where d denotes the metric on Q. Then

$$\|f + tg\| - \|f\| \geq |f(s_0) + tg(s_0)| - |f(s_0)| = -f(s_0) - tg(s_0) + f(s_0) = td(s_0, s_1)$$

for all sufficiently small t. We conclude that

$$\liminf_{t \to 0+} \frac{\|f + tg\| - \|f\|}{t} \geq d(s_0, s_1) > 0. \tag{1.6}$$

However,

$$\|f + tg\| - \|f\| \geq |f(s_1) + tg(s_1, s_1)| - |f(s_1)| = 0$$

for all t, whereby we must have

$$\limsup_{t \to 0-} \frac{\|f + tg\| - \|f\|}{t} \leq 0. \tag{1.7}$$

The inequalities (1.6) and (1.7) show that the limit (1.2) cannot exist.

For the case $f(s_0) > 0$, we define $g(s) \hat{=} d(s, s_1)$, and give a similiar argument. □

We now state and prove the theorem of Banach for surjective isomerties on $C(Q)$ space.

Theorem 2.2 (Banach)**.** *If Q and K are compact metric spaces then for the spaces of real continuous functions $C(Q)$ and $C(K)$ to be linear isometrically isomorphic it is necessary and sufficient that Q and K be homeomorphic. In this case, a linear isometric isomorphism T from $C(Q)$ onto $C(K)$ must be given by*

$$Tf(t) = h(t) f(\phi(t)) \text{ for } t \in K, \tag{1.8}$$

where ϕ is a homeomorphism from K onto Q, and h is a real valued unimodular function on K.

Proof. It is easy to see that if ϕ is a homeomorphsim from K onto Q, then a transformation U defined by (1.8) is an isometric isomorphism of $C(Q)$ onto $C(K)$ and thus the sufficiency of the condition is clear.

For the necessity, let T be a linear isometry from $C(Q)$ onto $C(K)$, suppose $s_0 \in Q$, and let $f \in C(Q)$ be such that $|f(s_0)| > |f(s)|$ for all $s_0 \neq s \in Q$. By Lemma 2.1,

$$\lim_{r \to 0} \frac{\|f + rg\| - \|f\|}{r} = g(s_0) \operatorname{sgn} f(s_0)$$

must exist for every $g \in C(Q)$. Since T is an isometry,

$$g(s_0) \operatorname{sgn} f(s_0) = \lim_{r \to 0} \frac{\|f + rg\| - \|f\|}{r} = \lim_{r \to 0} \frac{\|Tf + rTg\| - \|Tf\|}{r}. \tag{1.9}$$

Then we apply Lemma 2.1 again to conclude that there is some $t_0 \in K$ such that

$$|Tf(t_0)| > |Tf(t)| \text{ for } \forall t_0 \neq t \in K.$$

(It is important to note here that we can apply the lemma since Tg runs through all of $C(K)$.) Furthermore, we may conclude from (1.9) and (1.3) that

$$g(s_0) \operatorname{sgn} f(s_0) = Tg(t_0) \operatorname{sgn} Tf(t_0). \tag{1.10}$$

If we let $h(t_0) \hat{=} \operatorname{sgn}(f(s_0)Tf(t_0))$, then $|h(t_0)| = 1$ and we define

$$Ug(t_0) \hat{=} h(t_0)g(s_0), \quad \text{for each } g \in C(Q). \tag{1.11}$$

Let us define ψ from Q to K by $\psi(s_0) = t_0$.

For well-defineness, assume that $f \neq \tilde{f} \in C(Q)$ satisfy $\|f\| = |f(s_0)| = |\tilde{f}(s_0)| = \|\tilde{f}\|$ and (1.1). Without loss of generalization, suppose $f(s_0) = \tilde{f}(s_0)$. Suppose $\|Tf\| = |Tf(t_0)| = \|T\tilde{f}\| = |T\tilde{f}(\tilde{t}_0)|$, If $t_0 \neq \tilde{t}_0$, $\left\|\frac{1}{2}(f + \tilde{f})\right\| = \|f\|$, but $\left\|\frac{1}{2}(Tf + T\tilde{f})\right\| < \|Tf\|$. It leads to contradiction with that T is an isometry. So ψ is also well-defined.

Then we are going to show that $U : C(Q) \to C(K)$ is well-defined. ψ is injective, for if $\psi(s_1) = \psi(s_2) = t_0$, then by (1.10) we have $|g(s_1)| = |g(s_2)|$ for $\forall g \in C(Q)$ since

$$|g(s_1)| = |Tg(t_0)| = |g(s_2)|$$

and so $s_1 = s_2$.

To see that ψ is surjective, let $t_0 \in K$ and defined q on K by

$$q(t) = \frac{1}{1 + d(t, t_0)},$$

where d denotes the metric on K. If $f = T^{-1}q$, then by (1.10)

$$|f(s)| = |q(\psi(t))| = \frac{1}{1 + d(\psi(s), t_0)}$$

for each $s \in Q$. Since $\|f\| = \|q\| = 1$, there exists $s_0 \in Q$ such that $|f(s_0)| = 1$. Therefore, $\frac{1}{1+d(\psi(s_0),t_0)} = 1$, which implies that $t_0 = \psi(s_0)$.

Then we are going to prove the well-defineness of $h(t_0)$, and hence the well-defineness of U. As what we have done before, assume that $f \neq \tilde{f} \in C(Q)$ satisfy $\|f\| = |f(s_0)| = |\tilde{f}(s_0)| = \|\tilde{f}\|$ and (1.1). Without loss of generalization, suppose $f(s_0) = \tilde{f}(s_0)$. Then if $\operatorname{sgn} Tf(t_0) \neq \operatorname{sgn} T\tilde{f}(t_0)$, then $\left\|\frac{1}{2}(f + \tilde{f})\right\| = \|f\|$, but $\left\|\frac{1}{2}(Tf + T\tilde{f})\right\| < \|Tf\|$. It leads to contradiction with that T is an isometry. So U is well-defined.

Finally, suppose that $\{s_n\}$ is a sequence in Q converging to $s_0 \in Q$, $g \in C(K)$ and $Tf = g$. Since $|g(\psi(s_n))| = |f(s_n)|$ for each n and f is continuous, we have $|f(s_n)| \to |f(s_0)|$ so that $|g(\psi(s_n))| \to |g(\psi(s_0))|$ for every $g \in C(K)$. By choosing g defined by $g(t) \hat{=} d(t, \psi(s_0))$, we obtain

$$d(\psi(s_n), \psi(s_0)) = |g(\psi(s_n))| \to |g(\psi(s_0))| = 0,$$

from which we conclude that $\psi(s_n) \to \psi(s_0)$. This shows that ψ is continuous. Since Q is compact and K is Hausdorff, ψ must be a homeomorphism. Hence Q and K are homeomorphic and if we let $\phi = \psi^{-1}$, we get from (1.11) the characterization of U given by (1.8). □

Cor 2.3 (Banach). *If T is a surjective linear isometry on $C(Q)$ where Q is compact and metric, then*

$$Tf(t) = h(t)f(\phi(t)),$$

where $|h(t)| = 1$ and ϕ is a homeomorphism of Q onto itself.

2.2 On compact Hausdorff space

Now we shall generalize the theorem established by Banach in the case of separable compact Hausdorff space (compact metric space). The proof of Stone is necessarily somewhat different from the proof of Banach.

Theorem 2.4 (Stone)**.** *If Q and Q^* are compact Hausdorff spaces and M and M^* are corresponding continuous real-valued-function-rings. Then the existence of an isometric correspondence $f \to Uf =: f^*$ between M and M^* is equivalent to the existence of a topological equivalence $r \to \rho(r) =: r^*$ between Q and Q^*, the two correspondences being connected by the relations*

$$f(r) = \phi^*(r^*)\left[f^*(r^*) - \theta^*(r^*)\right],$$

$$Uf = f^*,\ \rho(r) = r^*,\ U0 = \theta^*,\ U1 = \phi^*,$$

where $|\phi| \equiv 1$. If the relations $U0 = 0, U1 = 1$ or, equivalently, the relations $\theta^* = 0, \phi^* = 1$ are satisfied, then the correspondence $f \to Uf = f^*$ is an analytical isomorphism between M and M^*.

An isometric correspondence U means one with the property$\|Uf - Ug\| = \|f - g\|$. It is evident that such a correspondence is one-to-one and has an isometric inverse U^{-1}. An analytical isomorphism of C^*-algebra is a C^*-isomorphism which maps one distinguished subalgebra, the analytic subalgebra, onto another.

Proof. The part of sufficiency is easy. When Q and Q^* are homeomorphic, and the homeomorphism is given by ρ. ϕ^* and θ^* is given arbitrarily (satisfying $|\phi| \equiv 1$). We define U by

$$f^*(r^*) = Uf(r^*) := f(\rho^{-1}(r^*))/\phi(r^*) + \theta(r^*).$$

It is easily vertified that U carries M isometrically onto M^*. With the special choice $\phi^* = 1, \theta^* = 0$, it is evident that U determines an analytical isomorphism between M and M^*. We may remark that if U defines such an isomorphism, the necessary relations $U0 = 0, U1 = 1$ imply $\theta^* = 0, \phi^* = 1$.

It will take more effort to prove the part of necessity.

(1)First we claim that the isometry must be kind of 'linear'.

When the isometric correspondence U is given, we define $\theta^* = U0$ and determine a new correspondence V by the relations

$$f \to Vf := Uf - \theta^* = f^* - \theta^*.$$

It is evident that V carries M isometrically onto M^* and that V has the additional property $V0 = 0$. A theorem of Marzur and Ulam (see Lemma 2.8) shows that V is a linear conrespondence, satisfying the relation

$$V(\alpha f + \beta g) = \alpha Vf + \beta Vg.$$

(2)We now construct the topological equivalence $\rho(r)$ in terms of the correspondence V.

If $r \in Q$, we define $M(r) := \{f \in M : |f(r)| = \|f\|\}$. It is evident that $M(r) \supseteq \mathbb{R}$. Also, if $f_1, \cdots, f_n \in M(r)$, the function $g := \sum_{\nu=1}^n f_\nu \operatorname{sgn} f_\nu(r) \in M(r)$ and satisfies the relation $\|g\| = \sum_{\nu=1}^n \|f_\nu\|$, as we infer from the inequalities

$$\|g\| \leq \sum_{\nu=1}^{n} \|f_\nu\| = \sum_{\nu=1}^{n} |f_\nu(r)| = g(r) \leq \|g\|. \tag{1.12}$$

We can now let $\mathscr{F}_f^* := \{r^* : |f^*(r^*)| = \|f^*\|\}$, where $f^* = Vf$, and $f \in M(r)$. Since f^* is continuous, $\mathscr{F}_f^*$ is closed. We claim that $\cap_{f\in M(r)}\mathscr{F}_f^* \neq \varnothing$. Since Q^* is compact, it suffices to prove that any intersection of a finite number of sets $\mathscr{F}_f^*$ is non-void. Consider $\mathscr{F}_1^*, \cdots, \mathscr{F}_n^*$, which are corresponding to $f_1, \cdots, f_n \in M(r)$. Consider g defined above and its correspondnece

$$g^* := Vg = \sum_{\nu=1}^{n} V\left(f_\nu \operatorname{sgn} f_\nu(r)\right) = \sum_{\nu=1}^{n} (Vf_\nu) \operatorname{sgn} f_\nu(r) = \sum_{\nu=1}^{n} f_\nu^* \operatorname{sgn} f_\nu(r). \tag{1.13}$$

Since Q^* is compact, there exists a point $p^* \in Q^*$ such that $|g^*(p^*)| = \|g^*\|$. We now observe the relations

$$\|g^*\| = |g^*(p^*)| \leq \sum_{\nu=1}^{n} |f_\nu^*(p^*)| \leq \sum_{\nu=1}^{n} \|f_\nu^*\| = \sum_{\nu=1}^{n} \|f_\nu\| = \|g\| = \|g^*\|. \tag{1.14}$$

Then we must have $|f_\nu^*(p^*)| = \|f_\nu^*\|$ for $\forall \nu = 1, \cdots, n$. Hence the point p^* is common to $\mathscr{F}_1^*, \cdots, \mathscr{F}_n^*$. Then we have $\cap_{f\in M(r)}\mathscr{F}_f^* \neq \varnothing$.

Let $r^* \in \cap_{f\in M(r)}\mathscr{F}_f^*$. It is evident that $f \in M(r)$ and $f^* = Vf$ imply $f^* \in M^*(r^*)$. In other words, that V carries $M(r)$ into a subclass of $M^*(r^*)$. By symmetry, the inverse correspondence V^{-1} carries $M^*(r^*)$ into some class $M(p)$. The inclusion relation $M(r) \supseteq M(p)$ is obvious. And it implies $p = r$. In fact, if $p \neq r$, since Q is compact Hausdorff, hance normal (T4), according to Urysohn Lemma, there exists a function $f \in M$ such that $f(p) = 0, f(r) = 1, 0 \leq f \leq 1$, then $f \in M(r)$ but $f \notin M(p)$. We can now infer that V carries $M(r)$ into $M^*(r^*)$, and, further, that V carries the family of all classes $M(r)$ one-to-one onto the famliy of all classes $M^*(r^*)$.

This correspondence between the classes $M(r)$ and $M^*(r^*)$ determines a one-to-one correspondence $r \leftrightarrow r^* =: \rho(r)$ between Q and Q^*. Now we are going to prove that the latter correspondence is a homeomorphism. From the structure of $\rho(r)$, it is seen that $A_f := \{r \in Q : |f(r)| = \|f\|\}$ and $A_{f^*}^* := \{r^* \in Q^* : |f^*(r^*)| = \|f^*\|\}$ correspond under the correspondence $r \leftrightarrow r^* := \rho(r)$ between Q and Q^* when $f^* = Vf$, since

$$r \in A_f \iff f \in M(r) \iff f^* \in M^*(r^*) \iff r^* \in A_{f^*}^*.$$

Then the complementary sets $Q \setminus A_f = \{r \in Q : |f(r)| < \|f\|\}$ and $Q^* \setminus A_{f^*}^* = \{r^* \in Q^* : |f(r^*)| < \|f^*\|\}$ correspond likewise.

(3)The correspondence we get is homeomorphism.

It is sufficient for us to prove that the latter sets constitute bases for the respective space Q and Q^*. Since the same discussion applies to both Q and Q^*, we may consider the space Q alone. Since Q is a compact Hausdorff space, hence a normal space. Suppose that there are more than one point in Q. If $L \subsetneq Q$ is a non-empty open set, then $K := Q \setminus L \subsetneq Q$ is a non-empty closed set. For $\forall r_0 \in L$, accroding to Urysohn Lemma, there exists $g \in M$ such that $g(r_0) = 0, g(K) = 1, 0 \leq g \leq 1$. So $r_0 \in \{r \in Q : |g(r)| < \|g\|\} \subseteq Q \setminus K = L$. And since Q is Hausdorff, a point in Q is a closed set. For $\forall r_0 \in Q$, choose $r_0 \neq r_1 \in Q$, then consider $L = Q \setminus \{r_1\}$, repeating the argument above, we can get

$g \in M$ such that $r_0 \in Q \setminus A_g$. In other words, $\cup_{g\in M}(Q \setminus A_g) = Q$. So $\{Q \setminus A_g\}_{g\in M}$ constitude bases for Q.

(4)The isometry and the homeomorphism have the relation:

$$f(r) = \phi^*(r^*)[f^*(r^*) - \theta^*(r^*)].$$

We define ϕ^* as the function $V1 \in M^*$. Since $\|\phi^*\| = \|f\| = \|1\| = 1$ and $A_1 = \{r \in Q : |f(r)| = 1\} = Q$, we conclude that $A^*_{\phi^*} = \{r^* \in Q^* : |\phi^*(r^*)| = 1\} = Q^*$, hence $|\phi| \equiv 1$. The correspondence W defined as

$$Wf := \phi^* Vf \tag{1.15}$$

therefore has the properties $|Wf| = |Vf|, \|Wf\| = \|Vf\| = \|f\|, W0 = 0, W1 = \phi^*\phi^* = 1$, and $W(\alpha f + \beta g) = \alpha Wf + \beta Wg$. Since the first two properties imply $A^*_{Vf} = A^*_{Wf}$, we see that the construction of the preceding paragraph leads to the same topological equivalencn ρ if we start with W rather than with V.

In terms of W, the relation between U, ϕ^*, θ^* and ρ which we wish to establish assumes the equivalent but simpler form $f(r) = f^*(r^*)$, where $f^* = Wf, r^* = \rho(r)$. As a first step in proving this relation, we show that $f \geq 0$ implies $Wf \geq 0$. If α and β are the minimum and maximun, respectively, of the function f. The relation $0 \leq \alpha \leq \beta$ imply that the function $g = \beta - f \geq 0$ has the number $\beta - \alpha$ as its maximum. Hence $\|g\| = \beta - \alpha$. We now have

$$f^* = Wf = W\beta + Wf - W\beta = \beta - Wg \geq \beta - \|Wg\| = \beta - \|g\| = \alpha \geq 0, \tag{1.16}$$

then we obtian the desired result. As a second step, we prove that $W|f| = |Wf|$. Since $|f| - f \geq 0$, we have $W|f| - Wf = W(|f| - f) \geq 0$, hence $W|f| \geq Wf$. Similarly, $|f| + f \geq 0$ implies $-W|f| \leq Wf$. We therefore conclude that $W|f| \geq |Wf|$. By symmetry, $W^{-1}|Wf| \geq |W^{-1}Wf| = |f|$. Since $W^{-1}|Wf| - |f| \geq 0$, $|Wf| - W|f| = W(W^{-1}|Wf| - |f|) \geq 0$, hence $|Wf| \geq W|f|$. So we conclude that $|Wf| = W|f|$.

We are now in a position to complete our proof. Let α be the value of f at a fixed point $r \in Q$ and let β be maximun of the function $|f - \alpha|$. Then the function $g = \beta - |f - \alpha| \geq 0$ belongs to M and has a maximal β at point r. Hence $\|g\| = \beta$. Since $r \in A_g = \{p \in Q : |g(p)| = \|g\|\}$, we see that $r^* = \rho(r) \in A^*_{g^*} = \{p^* \in Q^* : |g^*(p^*)| = \|g^*\|\}$ where $g^* = Wg$.

$$g^* = Wg = W(\beta - |f - \alpha|) = \beta - W|f - \alpha| = \beta - |Wf - W\alpha| = \beta - |f^* - \alpha| \tag{1.17}$$

$$|g^*| = |Wg| = W|g| = Wg = g^*, \ \|g^*\| = \|g\| = \beta \tag{1.18}$$

Hence we see that

$$\beta - |f^*(r^*) - \alpha| = g^*(r^*) = |g^*(r^*)| = \|g^*\| = \beta. \tag{1.19}$$

So $f^*(r^*) = \alpha = f(r)$. This completes the demonstration. □

Now we will go back to prove the lemma we used when proving the theorem above.

Definition 2.1 (center of a pair of points)**.** *If x, y are elements of a normed linear space (n.l.s.)$\mathcal{X}$. Let $H_1 = H_1(x, y)$ denote the set of elements $u \in \mathcal{X}$ such that $\|x - u\| = \|y - u\| = \frac{1}{2}\|x - y\|$. For $n = 2, 3, \cdots$, let H_n be the set of $u \in H_{n-1}$ so that $\|u - v\| \leq \frac{1}{2}\delta(H_{n-1})$ for $\forall v \in H_{n-1}$. Here $\delta(H_{n-1})$*

denotes the diameter of H_{n-1} which is, of course, the supremum of the distances between pairs of its elements. Clearly, $\delta(H_n) \le \frac{1}{2^{n-1}}\|x-y\|$ for each n. Hence, the intersection of H_n is either empty or consists of exactly one elements which is called the (metric) center of pair x, y.

Lemma 2.5. *If x, y are elements of a n.l.s. $\mathscr{X}$, then $\frac{1}{2}(x+y)$ is the center of the pair x, y.*

Proof. For each $u \in \mathscr{X}$, let $\tilde{u} = x + y - u$. If $u \in H_1(x, y)$, then $\|\tilde{u} - x\| = \|y - u\|, \|\tilde{u} - y\| = \|x - u\|$, so

$$\|\tilde{u} - x\| = \|\tilde{u} - y\| = \frac{1}{2}\|x - y\|,$$

because $u \in H_1$. Assume that $\tilde{u} \in H_{n-1}$ whenever $u \in H_{n-1}$ and let $u \in H_n$. If $v \in H_{n-1}$, we have

$$\|\tilde{u} - v\| = \|(x + y - v) - u\| = \|\tilde{v} - u\| \le \frac{1}{2}\delta(H_{n-1}),$$

since $\tilde{v} \in H_{n-1}$ and $u \in H_n$. Therefore, $\tilde{u} \in H_n$ as well. By induction, for each positive integer n, $\tilde{u} \in H_n$ whenever $u \in H_n$.

Next we show by induction that $z = \frac{1}{2}(x+y) \in H_n$ for each n. First we see that $z \in H_1$ since $\|z - x\| = \frac{1}{2}\|x - y\|$. Assume that $z \in H_{n-1}$ and $u \in H_{n-1}$. Thus $\tilde{u} \in H_{n-1}$. Then

$$2\|z - u\| = \|x + y - 2u\| = \|\tilde{u} - u\| \le \delta(H_{n-1}).$$

Hence, $\|z - u\| \le \frac{1}{2}\delta(H_{n-1})$ and we have $z \in H_n$. The conclusion is that $z \in \cap_{n=1}^{\infty} H_n$ and so it is the center of x, y. □

If T is an isometry from a n.l.s $\mathscr{X}$ into a n.l.s. $\mathscr{Y}$, then T obviously maps $H_1(x, y)$ into $H_1(Tx, Ty)$, since

$$\|Tu - Tx\| = \|u - x\| = \frac{1}{2}\|x - y\| = \frac{1}{2}\|Tx - Ty\|,$$

and similarly for $\|Tu - Ty\|$. This requires no linearity nor surjectivity for T.

Lemma 2.6. *If T is a surjective isometry from $\mathscr{X}$ to $\mathscr{Y}$, then T maps the center of any pair x, y in $\mathscr{X}$ to the center of the pair Tx, Ty in $\mathscr{Y}$.*

Proof. Since T is a surjective isometry, it is clear that $T(H_1(x, y)) = H_1(Tx, Ty)$ and $\delta(H_1(x, y)) = \delta(H_1(Tx, T, y))$. If we assume this holds for $H_{n-1}(Tx, Ty)$, then for $u \in H_n(x, y)$ and $w \in H_{n-1}(Tx, Ty)$, we have $w = Tv$ for some $v \in H_{n-1}(x, y)$ while

$$\|Tu - Tv\| = \|u - v\| \le \frac{1}{2}\delta(H_{n-1}(x, y)) = \frac{1}{2}\delta(H_{n-1}(Tx, Ty)).$$

Therefore, $Tu \in H_n(Tx, Ty)$. Similarly, if $w \in H_n(Tx, Ty)$, then $w = Tu$ for some $u \in H_n(x, y)$. Thus by induction, $T(H_n(x, y)) = H_n(Tx, Ty)$ for every positive integer n. By Lemma 2.5, $\frac{1}{2}(x+y) \in H_n(x, y)$ for every n and it follows from the statement above that $T\left(\frac{1}{2}(x+y)\right) \in H_n(Tx, Ty)$ for every n. Hence $T\left(\frac{1}{2}(x+y)\right)$ is the unique element of $\cap_{n=1}^{\infty} H_n(Tx, Ty)$ which is $\frac{1}{2}(Tx, Ty)$ by Lemma 2.5 again. □

Lemma 2.7. *If T is an isometry from a n.l.s. $\mathscr{X}$ onto a n.l.s. $\mathscr{Y}$ (real or complex), then*

(i) $T(x+y) = T(x) + T(y) - Y(0)$,

(ii) $T(sx) = sT(x) + (1-s)T(0)$ for all real numbers s.

Proof. If $x \in \mathscr{X}$, then from Lemma 2.6, we have

$$T(x) = T\left(\frac{1}{2}(2x+0)\right) = \frac{1}{2}(T(2x)+T(0)),$$

so that

$$T(2x) = 2T(x) - T(0). \tag{1.20}$$

Now applying Lemma 2.6 again, $T(x+y) = T\left(\frac{1}{2}(2x+2y)\right) = \frac{1}{2}(T(2x)+T(2y))$ and a double application of (1.20) yields (i) above.

Using (i) it is easy to show by induction that (ii) holds for all positive integers. Since

$$T(0) = T(x+(-x)) = T(x) + T(-x) - T(0),$$

we conclude that (ii) holds for $s = -1$ and therefore for any integer n. Upon applying (ii) to $T(x) = T(n\frac{x}{n})$ we get that (ii) holds for all rationals and the extention to real follows from the continuity of T. □

Lemma 2.8. *[Mazur-Ulam] If T is an isometry from a n.l.s. $\mathscr{X}$ onto a n.l.s. $\mathscr{Y}$, and if $T(0) = 0$, then T is real linear.*

Proof. The proof is immediate from Lemma 2.7. □

2.3 Notes and remarks

Theorem 2.2 is the original version of what has since been called the Banach-Stone Theorem. Stone proved the Theorem in the case where the space Q and K are assumed only to be compact and Hausdorff. If $C(Q,E)$ denotes the space of continuous functions from a compact Hausdorff space Q into a Banach space E, then E is said to have the **Banach-Stone Property** if the conclusions of Theorem 2.2 hold for isometric isomorphisms from $C(Q,E)$ to $C(K,E)$. Thus Stone showed that the real numbers have the Banach-Stone Property. Jersion was the first th consider the question for more general Banach space E.

Theorem 2.2 have been proved in a number of ways and by all of the principal methods that are used to find isometries in Banach spaces. Banach's method of using the directional derivative of the norm to identify peak points of functions has not appeared often in the works of other authors. Sundaresan has proved a generalization of Lemma 2.1 and used it in a proof of a Banach-Stone Theorem for spaces $C(Q,E)$, but not in a direct way to find the homeomorphism as Banach did. Okikiolu used differentiation of norms to determine isometries in L^p-space. W.Werner used a differentiability property to characterize isometries on C^*-algebras.

Reference

[1] Stone, Marshall Harvey. 'Applications of the theory of Boolean rings to general topology.' Transactions of the American Mathematical Society 41 (1937): 469-473.
[2] Richard J. Fleming, James E. Jamison. 'Isometries on Banach Spaces: function spaces': 3-8, 19-20
[3] Zuoqin Wang, Topology (Available on his website)

微分方程在期权定价问题中的应用

严浩

摘 要

期权是一种重要的金融衍生证券, 其定价问题是当今金融业发展的重要基础. 本文会先介绍期权定价问题的引入动机与基本的期权类型, 然后分别使用微分方程求解理论与数值计算方法给出定价问题的解. 最后简要介绍近代期权定价问题的模型与求解技术的发展情况.

1 问题的引入——什么是期权定价问题?

期权作为一种衍生证券, 它的定价决定于原生资产价格的变化. 由于原生资产是一种风险资产, 因此它的价格是随机的. 由此产生的期权价格变化也是随机的. 但是一旦原生资产价格确定下来, 那么作为它的衍生证券 (期权) 的价格也将随之确定. 这就是说, 若在 t 时刻原生资产价格为 S_t, 期权价格为 V_t, 则存在函数 $V(S,t)$ 使得 $V_t = V(S_t,t)$. 这里 $V(S,t)$ 是一个确定的二元函数, 我们的任务就是通过建立偏微分方程模型去确定这个函数. 特别地, 代入 $t=0$ 即得期权金 $p = V(S_0, 0)$.

换句话说, 假设你在时刻 $t=0$ 卖给某个人一个权利, 使得他在将来某个时刻 $t=T$ 以给定价格 K 去购买 (出售) 某种风险资产, K 的值与交易日当天风险资产的实际价值无关[1], 且持有者可自行决定行权与否[2], 这种权利叫做欧式看涨 (看跌) 期权. 那么为了保证你不受损失, 你售卖期权的价格应该是多少? 这就是所谓的欧式期权定价问题.

我们的行文思路如下: 利用 Black-Scholes 公式对欧式期权定价问题给出解析解, 并进一步讨论更复杂的美式期权求解理论, 借此引入有限差分法, 最终计算出美式期权定价.

2 欧式期权

2.1 基本假设

(a) 原生资产价格演化遵循几何 Brown 运动[3]

$$\frac{\mathrm{d}S_t}{S_t} = \mu \mathrm{d}t + \sigma \mathrm{d}W_t,$$

其中 μ、σ 是常数, $\mathrm{d}W_t$[4]是标准 Brown 运动, 满足 $\mathbb{E}[\mathrm{d}W_t]=0, \mathrm{Var}[\mathrm{d}W_t]=\mathrm{d}t$,

(b) 无风险利率 r 是常数,

(c) 原生资产不支付股息,

作者信息: 严浩, 2021 级数学科学学院, 邮箱: colin_yan@mail.ustc.edu.cn.

稿件信息: 2023 年 2 月 24 日来稿, 5 月 24 日修改, 5 月 26 日接受, 稿件编辑: 王乐达.

[1] "K 值与交易日风险资产实际价值无关" 这个性质解释了看涨与看跌这两个名字的由来——看涨即希望花小钱买进值钱的股票, 看跌即希望把不值钱的股票高价售卖.

[2] "持有者可决定行权与否" 这个性质解释了为什么我们要引入 f^+(细节见下文)——毕竟一个清醒的持期权者如果知道行权会亏损, 那么他就不会行权, 期权此时变成了一张废纸!

[3] 几何 Brown 运动是一种简单的 Ito 过程, 对于不熟悉 Ito 过程的读者, 只需要知道 Ito 过程可以通过 Ito 引理转化成较容易处理的 PDE(见下文).

[4] 在随机过程中, 我们一般用记号 X_t 表示随机变量 X 是 t 的函数, 而不用记号 $X(t)$, 不过读者可以认为这两个记号意义相同.

(d) 不支付交易费和税收,

(e) 无套利机会.[5]

2.2 问题与求解

设 $V=V(S,t)$ 是期权价格, 它在期权的到期日 $t=T$ 时

$$V(S,T)=\begin{cases}(S-K)^+ & (\text{看涨期权})\\ (K-S)^+ & (\text{看跌期权})\end{cases}, \text{其中} f^+ := \begin{cases} f\ , f>0 \\ 0\ , f\leqslant 0\end{cases}.$$

这里 K 是期权的敲定价 (不是期权金!). 问在期权的有效时间内它的价值?

为了解决这个问题, 我们先来推导 $V=V(S,t)$ 所满足的偏微分方程.

由基本假设,

$$\mathbb{E}[\mathrm{d}W_t]=0, \mathrm{Var}[\mathrm{d}W_t]=\mathrm{d}t=\mathbb{E}[(\mathrm{d}W_t)^2]-\mathbb{E}^2[\mathrm{d}W_t]=\mathbb{E}[(\mathrm{d}W_t)^2].$$

这个结果给我们一个不太严谨, 但十分重要的想法: 在不计高阶小量的情况下, 我们可以认为

$$(\mathrm{d}W_t)^2=\mathrm{d}t,$$

下面我们用这个条件证明十分重要的 Ito 公式, 进而导出我们需要的偏微分方程.

设随机变量 $V_t=V(S_t,t)$, V 是二元可微函数, 若随机过程 S_t 适合随机微分方程

$$\mathrm{d}S_t=\mu(S_t,t)\mathrm{d}t+\sigma(S_t,t)\mathrm{d}W_t,$$

则我们有如下的 Ito 公式:

$$\begin{aligned}\mathrm{d}V_t&=\left(\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2(S_t,t)\frac{\partial^2 V}{\partial S^2}\right)\mathrm{d}t+\frac{\partial V}{\partial S}\mathrm{d}S_t\\&=\left(\frac{\partial V}{\partial t}+\mu(S_t,t)\frac{\partial V}{\partial S}+\frac{1}{2}\sigma^2(S_t,t)\frac{\partial^2 V}{\partial S^2}\right)\mathrm{d}t+\sigma(S_t,t)\frac{\partial V}{\partial S}\mathrm{d}W_t.\end{aligned}$$

为了看清这一点, 对 V_t 作 Taylor 展开, 得到

$$\mathrm{d}V_t=\frac{\partial V}{\partial t}\mathrm{d}t+\frac{\partial V}{\partial S}\mathrm{d}S_t+\frac{1}{2}\frac{\partial^2 V}{\partial S^2}(\mathrm{d}S_t)^2+O(\mathrm{d}t\mathrm{d}S_t).$$

其中, 由基本假设,

$$\begin{aligned}(\mathrm{d}S_t)^2&=(\mu(S_t,t)\mathrm{d}t+\sigma(S_t,t)dW_t)^2\\&=\sigma^2(S_t,t)(dW_t)^2+2\mu\sigma\mathrm{d}t\mathrm{d}W_t+\mu^2\mathrm{d}t^2\\&=\sigma^2(S_t,t)\mathrm{d}t+o(\mathrm{d}t).\end{aligned}$$

将其带回原式, 略去高阶小量, 即可得到 Ito 公式.

我们注意到, Ito 公式中含有随机变量的微分, 也就是不确定项 $\mathrm{d}W_t$. 下面我们使用金融中的 $\Delta-$ 对冲技巧, 通过构造投资组合的方式消去不确定项, 从而得到可供计算用的 PDE.

$\Delta-$ 对冲技巧的主要想法是, 构造投资组合 $\Pi=V-\Delta S$, 其中 $\Delta=\Delta(t)$ 为原生资产份额, 选取适当的 Δ 使得在 $(t,t+\mathrm{d}t)$ 时段内, Π 无风险 (也就是说, 不含 $\mathrm{d}W_t$). 在这个组合的构造中, 原生资产份额 Δ 随着 t 时时刻刻变化, 这对应着实际生活中, 投资组合 Π 也是时时刻刻变化的.

[5]上述基本假设的合理性的完整推导、“标准 Brown 运动”的详细性质等细节需要随机过程等高阶课程, 在这里我们暂且接受这些条件.

设在时刻 t 有投资组合 Π, 并在时段 $(t, t+\mathrm{d}t)$ 时段内不改变份额 Δ, 那么由于 Π 是无风险的, 由基本假设, 在时刻 $t+\mathrm{d}t$, 投资组合的回报遵循无风险利率公式

$$\frac{\Pi_{t+dt}-\Pi_t}{\Pi_t}=r\mathrm{d}t,$$

即

$$\mathrm{d}V_t-\Delta\mathrm{d}S_t=r\Pi_t\mathrm{d}t=r(V_t-\Delta S_t)\mathrm{d}t.$$

应用 Ito 公式替换 $\mathrm{d}V_t$, 得到

$$\left(\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2S^2\frac{\partial^2V}{\partial S^2}+\mu S\frac{\partial V}{\partial S}-\Delta\mu S\right)\mathrm{d}t+\left(\sigma S\frac{\partial V}{\partial S}-\Delta\sigma S\right)\mathrm{d}W_t=r(V-\Delta S)\mathrm{d}t.$$

由于等式右端无风险, 因此左端随机项 $\mathrm{d}W_t$ 系数应为 0, 即选取

$$\Delta=\frac{\partial V}{\partial S}$$

代入 Δ, 等式两端约去 $\mathrm{d}t$, 我们就得到如下的 Black-Scholes 方程:

$$\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2S^2\frac{\partial^2V}{\partial S^2}+rS\frac{\partial V}{\partial S}-rV=0. \tag{1.1}$$

因此, 为了确定在合约有效期 $[0,T]$ 内期权的价值, 就要在区域 $\Sigma:\{0\leqslant S\leqslant\infty, 0\leqslant t\leqslant T\}$ 上求解定解问题

$$\begin{cases}\dfrac{\partial V}{\partial t}+\dfrac{1}{2}\sigma^2S^2\dfrac{\partial^2V}{\partial S^2}+rS\dfrac{\partial V}{\partial S}-rV=0 & (\Sigma)\\ V|_{t=T}=\begin{cases}(S-K)^+ & (\text{看涨期权})\\ (K-S)^+ & (\text{看跌期权})\end{cases}\end{cases}.$$

第一式是一个变系数抛物型 Euler 方程, 可通过换元转化成常系数方程; 另外, 该定解问题也是一个倒向定解问题. 特别地, 尽管直线段 $\{S=0, 0\leqslant t\leqslant T\}$ 是区域 Σ 的边界, 但由于 $S=0$ 是原方程的蜕化线, 故由 PDE 中的结论, 在 $S=0$ 上不必给值. 作变换

$$x=\ln S, \tau=T-t, V=u\mathrm{e}^{\alpha\tau+\beta x},\ \text{其中}\ \alpha=-r-\frac{1}{2\sigma^2}(r-\frac{\sigma^2}{2})^2, \beta=\frac{1}{2}-\frac{r}{\sigma^2}, \tag{1.2}$$

则原定解问题转化成我们熟悉的热方程

$$\begin{cases}\dfrac{\partial u}{\partial\tau}-\dfrac{\sigma^2}{2}\dfrac{\partial^2u}{\partial x^2}=0 & (\Sigma)\\ u|_{\tau=0}=\begin{cases}\mathrm{e}^{-\beta x}(e^x-K)^+ & (\text{看涨期权})\\ \mathrm{e}^{-\beta x}(K-e^x)^+ & (\text{看跌期权})\end{cases}\end{cases}.$$

以看涨期权初值条件为例, 套用热方程解的公式

$$u(x,\tau)=\int_{-\infty}^{+\infty}K(x-y,\tau)\varphi(y)\mathrm{d}y,$$

其中 $K(x-y,\tau)=\frac{1}{\sigma\sqrt{2\pi\tau}}\mathrm{e}^{\frac{-(x-y)^2}{2\sigma^2\tau}}$, $\varphi(y)$ 是初值, 可得

$$u(x,\tau)=\int_{\ln K}^{+\infty}\frac{1}{\sigma\sqrt{2\pi\tau}}\mathrm{e}^{\frac{-(x-y)^2}{2\sigma^2\tau}}[\mathrm{e}^{(1-\beta)y}-K\mathrm{e}^{-\beta y}]\mathrm{d}y.$$

代入 (1) 第三式, 并引入变换

$$\eta = x - \xi + \left(r - \frac{\sigma^2}{2}\tau\right), \omega = \frac{\eta + \sigma^2\tau}{\sigma\sqrt{\tau}}, N(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-\frac{\omega^2}{2}}\mathrm{d}\omega.$$

代入整理可得

$$\begin{aligned}
&V(x,\tau) = I_1 + I_2,\\
&I_1 = \mathrm{e}^x N\left(\frac{x - \ln K + (r + \frac{\sigma^2}{2}\tau)}{\sigma\sqrt{\tau}}\right),\\
&I_2 = -K\mathrm{e}^{-r\tau}\mathrm{e}^x N\left(\frac{x - \ln K + (r - \frac{\sigma^2}{2}\tau)}{\sigma\sqrt{\tau}}\right).
\end{aligned}$$

令

$$d_1 = \frac{\ln\frac{S}{K} + (r + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}, \quad d_2 = d_1 - \sigma\sqrt{T-t}.$$

并代入 (1) 第一式和第二式, 即可得到欧式看涨期权的定价公式

$$V(S,t) = SN(d_1) - K\mathrm{e}^{-r(T-t)}N(d_2).$$

给定初始时间的股价 S_0, 代入 $(S_0, 0)$ 即可得到期权金 (期权定价).

对于看跌期权, 我们也可以重复上述求解过程, 但我们在此采用一种更迅捷的方法.

运用如下的看涨——看跌期权的平价公式:

$$c_t + K\mathrm{e}^{-r(T-t)} = p_t + S_t,$$

其中 c_t 与 p_t 表示 t 时刻的看涨 (call)、看跌 (put) 期权价格, S_t 表示 t 时刻原生资产的价值. 这个公式给出了看涨、看跌期权价格的关系, 知道其中一个之后, 可以轻易算出另一个.

将先前得到的 $V(S,t)$ 代入 c_t, 我们立刻得到欧式看跌期权的定价公式

$$\begin{aligned}
p_t &= c_t + K\mathrm{e}^{-r(T-t)} - S_t\\
&= K\mathrm{e}^{-r(T-t)}[1 - N(d_2)] - S[1 - N(d_1)]\\
&= K\mathrm{e}^{-r(T-t)}N(-d_2) - SN(-d_1).
\end{aligned}$$

其中用到了性质 $1 - N(x) = N(-x)$.

上述公式可以应用在解决支付红利、两值期权与复合期权的情形, 区别在于需要对基本假设作小的修正; 求解流程为将方程变为能套用 B-S 公式的形式, 而非重复解方程; 以及在该基本假设下建立平价公式, 感兴趣的读者可以参考 [1].

3 美式期权

经过上面的介绍, 我们已经对“期权是什么”有了一个基本的认知. 下面我们介绍美式期权.

美式期权的不同之处在于它可以随时行权. 由于可以提前实施, 持有人将拥有比欧式期权更多的获利机会, 因此一般来说它比欧式期权更贵一些. 能否抓住有利时机, 适时地实施这张合约以获利, 是每一个美式期权持有者都必须考虑的问题.

数学上, 美式期权的定价问题是一个自由边界问题, 所谓自由边界, 就是这样一条需要确定的交界线 S_0, 它把区域 $\{0 \leqslant S \leqslant \infty, 0 \leqslant t \leqslant T\}$ 分成两个部分, 一部分是继续持有区域

$$\Sigma_1 = \{S_0 \leqslant S < \infty\},$$

另一部分是终止持有区域

$$\Sigma_2 = \{0 \leqslant S < S_0\},$$

这条自由边界在金融上称为最佳实施边界. 我们先讨论最简单的永久美式期权, 再讨论一般的美式期权.

3.1 永久美式期权

永久美式期权是一张没有终止期, 在生效后任何时间都可以实施的美式期权. 为确定起见, 下面以看跌期权为例.

永久美式看跌期权的定价问题可以重述为: 求解在 Σ_1 上的期权价格 $V(S)$ 与最佳实施边界 S_0.

下面直接给出相应的定解问题[6]:

$$\begin{cases} \dfrac{\sigma^2}{2} S^2 \dfrac{d^2 V}{dS^2} + rS\dfrac{dV}{dS} - rV = 0, \quad S_0 < S < \infty, \\ V(S_0) = K - S_0, \\ V(\infty) = 0. \end{cases}$$

设解形如 $V = S^\alpha$, 代入解得 $\alpha_1 = 1, \alpha_2 = -\frac{2r}{\sigma^2}$. 结合边界条件可得解

$$V(S; S_0) = \left(\frac{S_0}{S}\right)^{\frac{2r}{\sigma^2}} (K - S_0).$$

这个解依赖于实施边界 S_0, 下面寻找最佳实施边界 $S = S_0^*$, 使得

$$V(S; S_0^*) = \max_{0 \leqslant S_0 \leqslant K} V(S; S_0)$$

对 $V(S; S_0)$ 关于 S_0 求偏导寻找极值点 S_0^*, 解得 $S_0^* = \frac{2rK}{2r+\sigma^2}$ 此时期权价格最大, 为

$$V(S) = \left(\frac{2rK}{2r+\sigma^2}\right)^{\frac{2r+\sigma^2}{\sigma^2}} S^{-\frac{2r}{\sigma^2}}.$$

类似于上面的平价公式, 我们可以导出永久美式看涨期权与看跌期权的对称关系, 快速导出对应的看涨期权定价与最佳实施边界, 在此不再赘述.

除此之外, 上述问题还可以写成变分不等方程:

求函数 $V = V(S) \in C^1_{[0,\infty)}$, 并具有分段连续二阶微商, 使得在 $[0, \infty)$ 上适合

$$\begin{cases} \min\{\mathcal{L}_\infty V, V - (K - S)^+\} = 0, \quad 0 < S < \infty, \\ V(0) = K, V(\infty) = 0. \end{cases}$$

其中

$$\mathcal{L}_\infty = \frac{\sigma^2}{2} S^2 \frac{d^2}{dS^2} + rS\frac{d}{dS} - r.$$

为永久美式期权下的 B-S 算子. 这种形式不显含自由边界. 我们下面的讨论将会着重用到这种形式的方程.

直接法和变分不等方程法各有特色, 采用哪种形式取决于我们需要研究的问题与拟选取的方法, 感兴趣的读者可以参考 [1].

[6]得出方程的手法仍然是 Ito 公式加 $\Delta-$ 对冲技巧, 这不是我们的重点.

3.2 一般美式期权

按金融意义, 美式期权可以分为两部分: 欧式期权定价, 与由于合约引入提前实施条款而需要增付的期权金. 欧式期权定价部分只需应用 B-S 公式, 我们已经解决; 而增付的期权金与最佳实施边界的位置有关. 从数学上讲, 我们希望导出最佳实施边界 $S = S(t)$ 所满足的方程式.

一般美式期权的定价问题是一个抛物型方程的自由边界问题. 为考虑更一般的情形, 下面引入支付红利的条件, 这在数学上仅仅是修改了一个系数, 没有本质的困难. 记

$$\mathcal{L} = \frac{\partial}{\partial t} + \frac{\sigma^2}{2} S^2 \frac{\partial^2}{\partial S^2} + (r-q)S\frac{\partial}{\partial S} - r$$

为 B-S 算子, 其中 q 为红利率, 其余符号的意义与上文保持一致.

篇幅起见, 我们直接给出美式期权的分解公式:

设 $V(S,t)$ 是美式看跌期权定价, 则

$$V(S,t) = V_E(S,t) + e(S,t),$$

其中

$$V_E(S,t) = Ke^{-r(T-t)}N(-\hat{d}_2) - Se^{-q(T-t)}N(-\hat{d}_1)$$

为支付红利的欧式看跌期权定价,

$$e(S,t) = \int_t^T d\eta \int_0^{S(\eta)} (Kr - q\xi) G(S,t;\xi,\eta) d\xi$$

为提前实施期权金, $G(S,t;\xi,\eta)$ 为 B-S 方程的基本解, 满足

$$\begin{cases} \mathcal{L}V = 0 \\ V(S,\eta) = \delta(S-\xi). \end{cases} \quad 0 < S < \infty,\ 0 < \xi < \infty,\ 0 < t < \eta,\ \delta(x) \text{ 为 Dirac 函数.}$$

上述公式说明, 如果最佳实施边界 $S = S(t)$ 已知, 那么美式期权定价可由上述分解公式表出. 不幸的是, $S(t)$ 已没有显式表达式, 事实上, $S = S(t)$ 满足非线性第二类 Volterra 积分方程:

$$\begin{aligned} S(t) = K &+ S(t)e^{-q(T-t)} N\left(-\frac{-\ln\frac{S(t)}{K} + \beta_2(T-t)}{\sigma\sqrt{T-t}}\right) \\ &- Ke^{-r(T-t)} N\left(-\frac{-\ln\frac{S(t)}{K} + \beta_1(T-t)}{\sigma\sqrt{T-t}}\right) \\ &- Kr\int_t^T e^{-r(\eta-t)}\left[1 - N\left(\frac{-\ln\frac{S(t)}{S(\eta)} + \beta_1(\eta-t)}{\sigma\sqrt{\eta-t}}\right)\right] d\eta \\ &+ qS\int_t^T e^{-q(\eta-t)}\left[1 - N\left(\frac{-\ln\frac{S(t)}{S(\eta)} + \beta_2(\eta-t)}{\sigma\sqrt{\eta-t}}\right)\right] d\eta, \end{aligned}$$

其中

$$\beta_1 = r - q - \frac{\sigma^2}{2}, \quad \beta_2 = r - q + \frac{\sigma^2}{2}.$$

求解这个非线性积分是很困难的, 从而我们并不能完满地给出美式期权的价格. 问题在这里开始变得棘手了起来.

4 数值方法

从上面的介绍中我们可以看出, 除永久美式期权以外, 一般的解都是没有显式表达式的. 另一方面, 欧式期权的定价即使有显式表达式, 但是在当今计算机的使用已相当普及的情况下, 人们有时还是乐于使用数值方法, 特别对于一些复杂的期权定价问题, 如上文提到的复合期权, 数值方法更显得有很多优越性.

因此, 我们下面将介绍数值方法, 并以此为理论基础编写求解程序. 用于求解美式期权定价问题的数值方法主要有有限差分法、二叉树法、切片法等, 我们下面将主要介绍有限差分法.

4.1 有限差分法简介

为进行形象的介绍, 我们以熟悉的热传导方程为例讲解有限差分法的操作流程:

$$\begin{cases} \dfrac{\partial u}{\partial t} - a^2 \dfrac{\partial^2 u}{\partial x^2} = 0, & 0 \leqslant x < \infty,\ t > 0, \\ u(0,t) = g(t), & 0 \leqslant t \leqslant T, \\ u(x,0) = \varphi(x). & 0 \leqslant x < \infty. \end{cases}$$

Step1. 划定网格

我们知道, 上述 PDE 的解的图像是 $\mathbb{R}^3$ 中的一张曲面, 其定义域是 xOy 平面 (的一部分), 直观上, 我们可以把这个平面换成一张“网格纸”, 也就是进行定义域的离散化, 通过求解网格纸上的样本点, 近似地织出一张解曲面.

在上述方程中, 定义域为 $\{0 \leqslant x < \infty,\ 0 \leqslant t \leqslant T\}$, 以间距 Δx 等分半直线 $0 \leqslant x < \infty$, 以间距 Δt 等分线段 $\leqslant t \leqslant T$. 记分点为 (x_m, t_n), 其中

$$x_m = m\Delta x\ (0 \leqslant m < \infty), \quad t_n = n\Delta t, \quad 0 \leqslant n \leqslant N,\ N = \frac{T}{\Delta t}.$$

并记函数 $u(x,t)$ 在每一个网格点 (x_m, t_n) 上的值为

$$u_m^n = u(x_m, t_n), \quad m = 0, 1, \cdots;\ n = 0, 1, \cdots, N.$$

Step2. 方程离散化与初边值离散化

由数学分析的知识, 我们有如下估计[7]:

$$f'(x) = \frac{f(x+\Delta x) - f(x)}{\Delta x} + O(\Delta x) \triangleq \left(\frac{\Delta f}{\Delta x}\right)_f + O(\Delta x) \quad \text{(前差分格式)}$$

$$f'(x) = \frac{f(x) - f(x-\Delta x)}{\Delta x} + O(\Delta x) \triangleq \left(\frac{\Delta f}{\Delta x}\right)_b + O(\Delta x) \quad \text{(后差分格式)}$$

$$f'(x) = \frac{f(x+\Delta x) - f(x-\Delta x)}{2\Delta x} + O(\Delta x^2) \triangleq \left(\frac{\Delta f}{\Delta x}\right)_c + O(\Delta x^2) \quad \text{(中心差分格式/Crank-Nicolson 格式)}$$

$$f''(x) = \frac{f(x+\Delta x) - 2f(x) + f(x-\Delta x)}{\Delta x^2} + O(\Delta x^2) \triangleq \left(\frac{\Delta^2 f}{\Delta x^2}\right)_c + O(\Delta x^2) \quad \text{(中心差分格式/Crank-Nicolson 格式)}$$

[7]事实上, 下面这三种一阶差分格式的计算性质差异很大: 前差分需要步长满足一定的限制关系时才具有稳定性; 后差分格式尽管具有无条件稳定性, 但是精度上不如中心差分格式. 这也说明了为什么中心差分格式得以被冠名——它的性质十分优良.

将热方程中的一阶偏导和二阶偏导用如上的差分替换掉, 并给出初边值的离散化形式, 就可以得到离散化的热方程.

显式差分格式

若用前差分替换一阶偏导, 我们将得到如下式子:

$$\begin{cases} \left(\dfrac{\Delta u}{\Delta t}\right)_f - a^2\left(\dfrac{\Delta^2 u}{\Delta x^2}\right)_c = 0, \\ u_0^n = g(t_n), \\ u_m^0 = \varphi(x_m). \end{cases}$$

即

$$\begin{cases} \left(\dfrac{u_m^{n+1} - u_m^n}{\Delta t}\right) - a^2\left(\dfrac{u_{m+1}^n - 2u_m^n + u_{m-1}^n}{\Delta x^2}\right) = 0, \\ u_0^n = g(t_n), \\ u_m^0 = \varphi(x_m). \end{cases}$$

整理可得

$$\begin{cases} u_m^{n+1} = (1-2\alpha)u_m^n + \alpha u_{m+1}^n + \alpha u_{m-1}^n, \\ u_0^n = g(t_n). \end{cases}$$

其中 $\alpha = \frac{\Delta t}{\Delta x^2}a^2$. 它的已知——未知决定关系如下图所示. 该算法可以直接地进行逐步求解, 故得名显式差分格式.

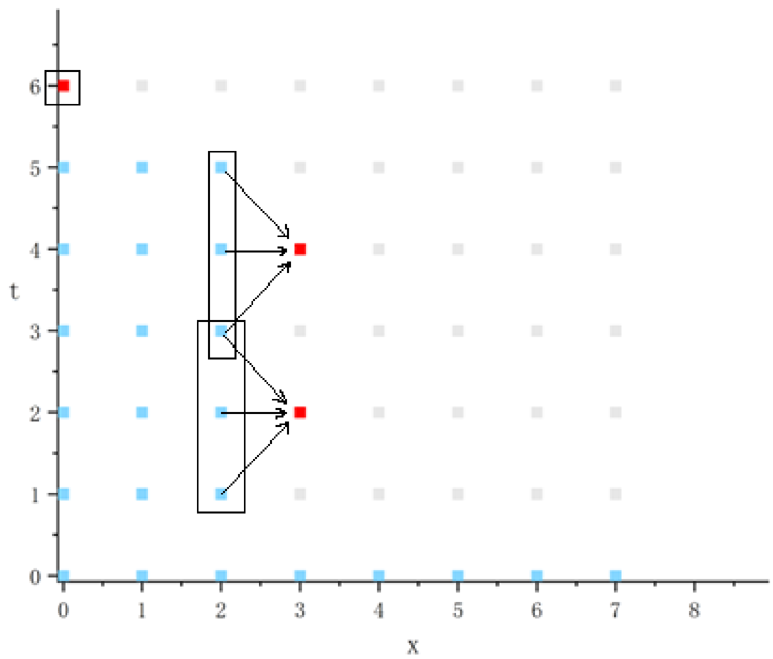


图 1.1: 显式差分格式下已知与未知数据点的决定关系.

其中蓝色数据点是已知的 (初边值/先前步骤计算得出), 红色数据点是接下来被计算出来的, 灰色数据点是尚未被计算出来的.

隐式差分格式

若用后差分替换一阶偏导, 我们将得到如下式子:

$$\begin{cases} \left(\dfrac{\Delta u}{\Delta t}\right)_b - a^2\left(\dfrac{\Delta^2 u}{\Delta x^2}\right)_c = 0, \\ u_0^n = g(t_n), \\ u_m^0 = \varphi(x_m). \end{cases}$$

即

$$\begin{cases} \left(\dfrac{u_m^{n+1} - u_m^n}{\Delta t}\right) - a^2\left(\dfrac{u_{m+1}^{n+1} - 2u_m^{n+1} + u_{m-1}^{n+1}}{\Delta x^2}\right) = 0, \\ u_0^n = g(t_n), \\ u_m^0 = \varphi(x_m). \end{cases}$$

整理可得

$$\begin{cases} (1+2\alpha)u_m^{n+1} - \alpha u_{m+1}^{n+1} - \alpha u_{m-1}^{n+1} = u_m^n, \\ u_0^{n+1} = g(t_{n+1}). \end{cases}$$

其中 $\alpha = \frac{\Delta t}{\Delta x^2}a^2$. 这里未知数变成了 u_m^{n+1}, 它的已知——未知决定关系如下图所示.

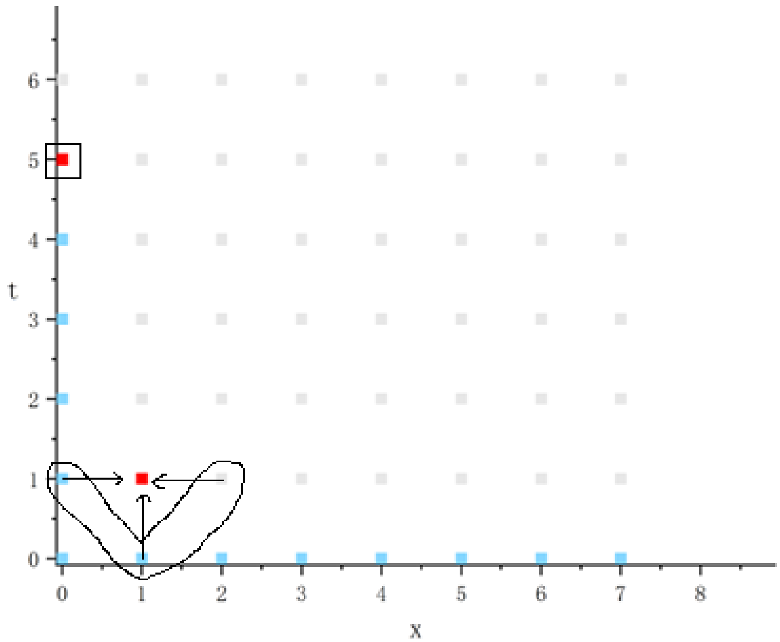


图 1.2: 隐式差分格式下已知与未知数据点的决定关系

左端坐标轴上的红色数据点仍然是可以轻易求出的, 但此时参与决定下方红色数据点 (u_m^{n+1}) 的数据点有一个是灰色的 (u_{m+1}^{n+1}, 未知), 故我们需要一个关于 u_{m+1}^{n+1} 的关系式, 但这时又会从右边引入新的未知点 $u_{m+2}^{n+1}\dots$

这也是隐式差分格式得名的原因: 无法通过单个方程求解, 而是要依靠线性方程组的联立来实现求解. 但是按照上面的过程来看, 未知数显然有无穷多个. 所以我们在求解隐式格式的时候通常需要在 $m = M$(M 充分大) 补充一个边界条件, 把它变成一个包含有限个 (($M-1$) 个) 未知数的线性方程组, 这涉及到解在无穷

远处的性态. 为简单起见, 我们假设 $\varphi(x)$ 收敛, 即

$$\lim_{x\to\infty}\varphi(x)=\varphi_\infty.$$

取 $u_M^{n+1}=\varphi_\infty$, 我们就得到了一个线性方程组, 其中

$$A\mathbf{u}^{(n+1)}=\mathbf{f}^{(n+1)},\ \mathbf{u}^{(n+1)}=\begin{bmatrix}u_1^{n+1}\\ \cdots\\ u_{M-1}^{n+1}\end{bmatrix},\ \mathbf{f}^{(n+1)}=\begin{bmatrix}f_1^{n+1}\\ \cdots\\ f_{M-1}^{n+1}\end{bmatrix},\ A=\begin{bmatrix}1+2\alpha & -\alpha & 0 & \cdots & 0\\ -\alpha & 1+2\alpha & -\alpha & \cdots & 0\\ 0 & -\alpha & 1+2\alpha & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & 1+2\alpha\end{bmatrix},$$

$$\alpha=a^2\frac{\Delta t}{\Delta x^2},\ f_1^{n+1}=u_1^n+\alpha g(t_{n+1}),\ f_i^{n+1}=u_i^n\ (2\leqslant i\leqslant M-2),\ f_{M-1}^{n+1}=u_{M-1}^n+\alpha\varphi_\infty.$$

由于 $\alpha>0$, 上文中 A 一定是一个对角优矩阵, 由线性代数的知识, 这个方程存在唯一解. 解这个方程组, 即可获得 $t=t_{n+1}$ 这一“层”中的 $u_m^{n+1},\forall m$.

另外, 需要指出的是, 由于数值计算中误差无法避免, 我们在做数值计算的时候还要考虑格式的稳定性, 显式格式计算简单, 但是需要控制 $\alpha<\frac{1}{2}$ 才能保证稳定性; 而隐式格式计算较复杂, 但对 α 无条件稳定.

4.2　美式期权定价的数值解

下面我们以显式差分格式为例, 求解美式期权定价. 我们先给出美式期权定价问题的的变分不等方程:

$$\begin{cases}\min\{-\mathcal{L}V,-(K-S)^+\}=0, & \Sigma,\\ V(S,T)=(K-S)^+, & 0\leqslant S<\infty,\\ V\to 0, & S\to\infty.\end{cases}$$

其中 $\mathcal{L}$ 为 B-S 算子, $\Sigma=\{0\leqslant S<\infty,0\leqslant t\leqslant T\}$.

Step1. 标准化

作变换

$$x=\ln\frac{S}{K},\ v(x,t)=\frac{V(S,t)}{K}.$$

则定解问题化为

$$\begin{cases}\min\{-\mathcal{L}_0v,v-(1-e^x)^+\}=0, & x\in\mathbb{R},\ 0\leqslant t\leqslant T,\\ v(x,T)=(1-e^x)^+, & x\in\mathbb{R},\\ v(+\infty,t)=0,v(-\infty,t)=1, & 0\leqslant t\leqslant T,\end{cases}$$

其中

$$\mathcal{L}_0=\frac{\partial}{\partial t}+\frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}+\left(r-\frac{\sigma^2}{2}\right)\frac{\partial}{\partial x}-r.$$

Step2. 划定网格与离散化

由于数值计算只能在有限区间上进行, 故我们需要对 x 所在的全直线作截断, 即只考虑 $[-L_1,L_2]\subset\mathbb{R}$ 上的求解, 相应地, 我们需要将端点处的值设为极限. 在带状区域 $\{-L_1\leqslant x\leqslant L_2,0\leqslant t\leqslant T\}$ 上划定网格

$$\pi=\{(n\Delta t,j\Delta x)|0\leqslant n\leqslant N,0\leqslant j\leqslant M\},$$
$$\Delta t=\frac{T}{N},\Delta x=\frac{L_1+L_2}{M}.$$

在每一个网格点上, 定义函数

$$v_j^n = v(j\Delta x, n\Delta t),$$
$$\varphi_j = \varphi(j\Delta x) = (1 - e^{j\Delta x})^+.$$

用前差分代替 v_t, 中心差分代替 v_x, 二阶中心差分代替 v_{xx}, 即

$$\left(\frac{\partial v}{\partial t}\right)_{n+1,j} = \frac{v_j^{n+1} - v_j^n}{\Delta t},$$
$$\left(\frac{\partial v}{\partial x}\right)_{n+1,j} = \frac{v_{j+1}^{n+1} - v_{j-1}^{n+1}}{2\Delta x},$$
$$\left(\frac{\partial^2 v}{\partial x^2}\right)_{n+1,j} = \frac{v_{j+1}^{n+1} - 2v_j^{n+1} + v_{j-1}^{n+1}}{\Delta x^2},$$

于是得到在格点 $v(j\Delta x, (n+1)\Delta t)$ 上的方程

$$\begin{cases} \min\left\{ -\dfrac{v_j^{n+1} - v_j^n}{\Delta t} - \dfrac{\sigma^2}{2}\dfrac{v_{j+1}^{n+1} - 2v_j^{n+1} + v_{j-1}^{n+1}}{\Delta x^2} - \left(r - \dfrac{\sigma^2}{2}\right)\dfrac{v_{j+1}^{n+1} - v_{j-1}^{n+1}}{2\Delta x} + rv_j^n, v_j^n - \varphi_j \right\} = 0, \\ v_j^N = \varphi_j. \quad 0 \leqslant n \leqslant N-1, 0 \leqslant j \leqslant M-1. \end{cases}$$

解得

$$v_j^n = \max\left\{\frac{1}{1 + r\Delta t}[(1-\omega)v_j^{n+1} + av_{j+1}^{n+1} + cv_{j-1}^{n+1}], \varphi_j\right\}, \quad 0 \leqslant n \leqslant N-1,\ 0 \leqslant j \leqslant M-1.$$

其中

$$\omega = \frac{\sigma^2\Delta t}{\Delta x^2},\ a = \frac{\omega}{2} - \frac{\Delta t}{2\Delta x}\left(r - \frac{\sigma^2}{2}\right),\ c = \omega - a.$$

逐步反向求解样本点, 直到 $n = 0$ 时, 就可得到美式期权的期权金. 注意, 对于上面的算法, 若满足

$$\omega \leqslant 1, \quad \frac{1}{\sigma^2}\left|r - q - \frac{\sigma^2}{2}\right|\Delta y \leqslant 1,$$

则显式差分格式收敛.

Step3. 程序与结果

用 MATLAB 编写模拟程序如下:

```
function [V x t]=
          ExplicitAmericanPut(k, L1, L2, T, phi, psi1, psi2, M, N, q, sigma, r)
%方程:同正文所示
%初值条件: v(x, T)=phi(x)
%边值条件: v(L1, t)=psi1(t). v(L2, t)=psi2(t)
%输出参数:V -解矩阵, 最后一行表示初值, 第一列和最后一列表示边值, 倒数第二行
    表示第2层......
%k -敲定价格
%x -空间变量
%t -时间变量
%输入参数:
%L1 -空间变量x的取值下限
%L2 -空间变量x的取值上限
%T -时间变量t的取值上限
%phi -初值条件, 定义为内联函数
```

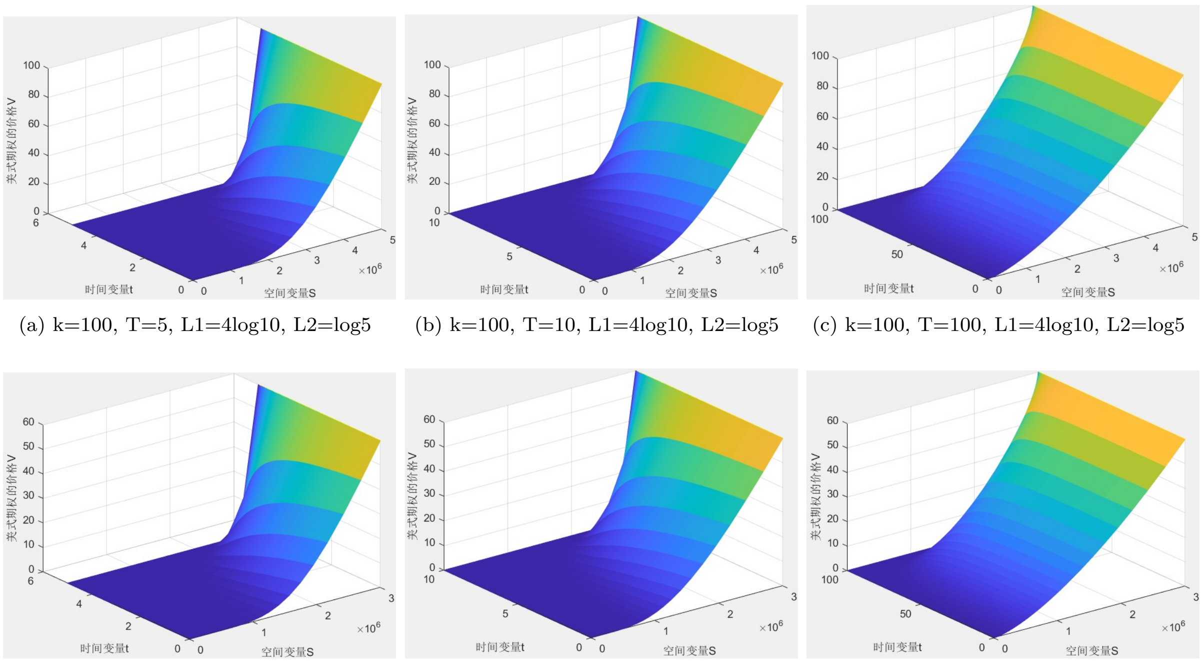
```
%psi1 -边值条件, 定义为内联函数
%psi2 -边值条件, 定义为内联函数
%M -沿x轴的等分区间数
%N -沿t 轴的等分区间数
%q -红利率(常数)
%sigma -波动率(常数)
%r -无风险利率(常数)
%计算步长
dx=(L1+L2)/M; %x的步长
dt=T/N; %t 的步长
x=(0:M)*dx;
t=(0:N)*dt;
omega=sigma^2*dt/dx/dx;
a=omega/2-dt/2/dx*(r-sigma^2/2);
c=omega-a;

if omega > 1
    disp('omega > 1, 不稳定')
end

if 1/(sigma*sigma)*abs(r-q-sigma*sigma/2)*dx > 1
    disp('第二个因子>1, 不稳定, 值为')
    disp(1/(sigma*sigma)*abs(r-q-sigma*sigma/2)*dx)
    disp(', dx值为')
    disp(dx)
end

%计算初值和边值
V=zeros(M+1, N+1);
for i=1:M+1
    V(i, N+1)=phi(x(i));
end
for j=1:N+1
    V(1, j)=psi1(t(j));
    V(M+1, j)=psi2(t(j));
end

%逐层求解
for j=1:N
    for i=2:M
        V(i, N+1-j)=max(1/(1+r*dt)*((1-omega)*V(i, N+2-j)+a*V(i+1, N+2-j)
                                        +c*V(i-1, N+2-j)), phi(x(i)));
    end
end
V=V';
%作出图形
mesh(k*exp(x), t, k*V);
title('古典显式格式, 变分不等方程解的图像')
```

```
xlabel('空间变量S')
ylabel('时间变量t')
zlabel('美式期权的价格V')
return;
%应用举例:在Matlab中打开该m文件, 然后在命令行中输入以下代码:
%q=0.07;r=0.03;sigma=0.2;M=100;N=512;k=100;T=5;L1=4*log(10);L2=log(5);
%phi=inline('max(0, 1-exp(x))');psi1=inline('0');psi2=inline('1')
%ExplicitAmericanPut(k, L1, L2, T, phi, psi1, psi2, M, N, q, sigma, r)
```

固定参数

$$q = 0.07, r = 0.03, \sigma = 0.2, M = 100, N = 512.$$

分别以如下参数条件作图:



(a) k=100, T=5, L1=4log10, L2=log5 (b) k=100, T=10, L1=4log10, L2=log5 (c) k=100, T=100, L1=4log10, L2=log5

(d) k=60, T=5, L1=4log10, L2=log5 (e) k=60, T=10, L1=4log10, L2=log5 (f) k=60, T=100, L1=4log10, L2=log5

图 1.3: 不同参数条件下的程序结果

4.3 结果分析

根据相关资料, 我们画出的定价曲面的形状是大致正确的, 并且实施边界曲线也具有凸性, 与现有的理论分析的结论一致 (见 [5]、[6]). 另一方面, 上述显示差分法虽然简单易行, 但收敛速度较慢, 且需要的空间截断应充分长, 导致计算量增大. 这些缺点导致了以上算法尚不能满足期权交易的实时性需求. 故对于更高效的算法, 应再进一步研究.

5 问题的延伸与总结

5.1 更先进的技术

CEV 模型

不变方差弹性模型 (CEV) 是由 Cox 和 Ross 在 1976 年提出的. 其对应的 PDE 为

$$\frac{\partial V}{\partial t}+\frac{\sigma^2}{2}S^{2\alpha}\frac{\partial^2 V}{\partial S^2}+(r-q)S\frac{\partial V}{\partial S}-rV=0,\quad \alpha>0,$$

其中 α 是弹性系数. 它是 Black-Scholes 模型的推广. 在这个模型中, 资产回报的波动率 $\sigma(S)=\sigma S^{\alpha-1}$ 是 S 的函数, 而不是常数.

Front-Fixing 方法与 PML 技巧

无论是 B-S 模型, 还是 CEV 模型, 在求解时都面临着大致相同的障碍, 可以概括为下面三点:

1. 求解区域左半部分的自由边界 $S_0(t)$ 为一未知曲线, 这样就使得求解的区域不规则, 数值求解较为困难;
2. 该问题求解区域的右半部分为一无界区域, 我们无法直接利用数值算法进行求解;
3. 一般的数值算法需同时计算自由边界 $S_0(t)$, 以及期权的价格 $V(S,t)$.

面对这种情况, 很多数学学者们都在探索更好的解决方法.

其中, Hongtao Yang 教授曾应用 Front-fixing 方法来求解自由边界问题.Front-Fixing 方法旨在将自由边界 (未知曲线) 变换成一条已知的直线, 进而转化为一维问题. 换言之, 该方法可将不规则的求解区域转换为规则的求解区域进行求解.Front-Fixing 方法的优点在于我们可以直接计算出自由边界, 且当自由边界为光滑或单调函数时, 该方法的精度会更高.

完全匹配层技巧 (PML) 最初是由 Berenger 在 1994 年提出的, Berenger 将此方法应用于麦克斯韦方程, 用来吸收无反射的电磁波, 从而使得截断误差更小.Lantos 也利用完全匹配层技巧 (PML) 来求解热传导方程. 这些技术的广泛应用大大促进了期权定价的方法论的发展.

5.2 奇异期权

除了这些标准的期权, 还有一些相对比较复杂的一些条款的期权, 文献中称其为奇异期权, 常见的有:

- 亚式期权: 到期收益取决于标的资产的平均价格.
- 回望期权: 到期收益取决于标的资产价格的最值.
- 障碍期权: 到期收益取决于标的资产价格是否达到了某一阈值.
- 篮子期权: 到期收益取决于多种标的资产的价格.

尽管期权的形式多种多样, 但是在求解其定价问题时仍然离不开 B-S 方程, 微分方程的应用之广泛, 在此可见一斑!

5.3 总结

本文用分析方法与数值方法讨论并求解了几种常见期权的定价问题. 如果要对技术细节作深入讨论, 或者得到能用于金融市场的高效期权定价方案, 则需引入更多的随机微分方程与统计学知识, 以便进一步研究.

参考文献

[1] 姜礼尚 期权定价的数学模型和方法 (第二版)[M] 北京, 高等教育出版社, 1-191. (2008)
[2] 蒲兴成、张毅 随机微分方程及其在数理金融中的应用 [M] 北京, 科学出版社, 1-3. (2010)
[3] 王生喜 金融数学理论与方法 [M] 北京, 清华大学出版社, 1-53. (2015)
[4] 陆金甫、关治 偏微分方程数值解法 (第二版)[M] 北京, 清华大学出版社, 13-28, 82-84. (2004)
[5] 王智宇 求解美式期权定价的带完全匹配层有限差分法 [D] 吉林, 吉林大学数学研究所, 1-9. (2015)
[6] 张琪 几类美式期权定价问题的数值方法 [D] 吉林, 吉林大学数学研究所, 13-16. (2016)

流体自由边界问题 (三)

——Beale-Kato-Majda 爆破准则

章俊彦

摘 要

欧拉方程组描述了理想流体的运动. 不可压缩欧拉方程在全空间 $\mathbb{R}^2$ 中存在整体解, 但 $\mathbb{R}^3$ 中不可压欧拉方程的解可能产生有限时间爆破. 1984 年, Beale-Kato-Majda [5] 证明了一个爆破准则 (称作 "BKM 爆破准则"): $\mathbb{R}^d$ 中的不可压欧拉方程解的 $H^s(s > \frac{d}{2}+1)$ 范数爆破, 当且仅当旋度 $L_t^1 L_x^\infty$ 范数在该时刻产生爆破. 1993 年, Ferrari [12] 将该结果推广到带边的有界区域.

若研究更现实的模型, 则应考虑不可压欧拉方程的**自由边界演化问题**, 此时自由区域的边界本身以及边界上流体的切向速度也应被考虑进来. 那么我们提出如下几个问题: 能否对自由边界问题得到 BKM 型爆破准则? 该爆破准则能否在边界固定时退化为 [12] 的结果? 准则里判定爆破的量, 其正则性要求能否达到最优? 另一方面, 二维理想流体的旋度是守恒量, 若能证出 BKM 型爆破准则, 则说明二维自由边界问题的解爆破必定是边界上先产生奇性而不可能是内部先爆破. 遗憾的是, 这些问题至今并没有完美的答案. 本文将简要介绍 [5, 12] 的证明思路、自由边界问题最新的进展、以及作者关于自由边界问题爆破准则的思考.

考虑一个不可压缩理想流体在全空间 $\mathbb{R}^d$ 中运动, 其速度场为 u, 压力为 p, 密度为 1. 则 u, p 满足如下欧拉方程组

$$\begin{cases} (\partial_t + u\cdot\nabla)u = -\nabla p & \text{in } [0,T]\times\mathbb{R}^d \\ \nabla\cdot u = 0 & \text{in } [0,T]\times\mathbb{R}^d \\ u|_{t=0} = u_0(x) & \text{on } \{t=0\}\times\mathbb{R}^d. \end{cases} \tag{0.1}$$

不难证明, 若欧拉方程具有局部化的初值 (即初值几乎集中于一个紧集内), 则欧拉方程的解具有守恒的能量 $E(t) = \frac{1}{2}\int_{\mathbb{R}^d} |u(t,x)|^2\,\mathrm{d}x$. 据此, 研究欧拉方程的适定性应当选择 Sobolev 型的函数空间 (例如 H^s) 而非连续函数空间 $C^{k,\alpha}$. 容易证明全空间 $\mathbb{R}^d$ 中的欧拉方程在 Sobolev 空间 $H^s(\mathbb{R}^d)$, $(s > \frac{d}{2}+1)$ 中有局部适定性. 那么一个自然的问题是全空间中的欧拉方程是否有长时间解甚至是整体解?

本文中, 我们只考虑空间维数为 2、3 的情况, 因为我们生活在三维的现实世界, 更高的维数没有意义. 实际上, 二维欧拉方程的整体适定性是容易证明的, 但三维欧拉方程的整体解却在很长一段时间内没有解决. 直到 2020 年, Tarek Elgindi [11] 首次构造了一个三维欧拉方程的解, 其可以在有限时间内爆破, 进而否定了三维欧拉方程的整体解存在性. 那么, 一个自然的问题是, 能否导出一个欧拉方程的 "爆破准则", 进而给出解在 Sobolev 空间中爆破的充分必要条件? 实际上在 1984 年, 这个问题就由 Beale-Kato-Majda [5] 给出了答案.

作者信息: 章俊彦, 2013 级数学科学学院, 邮箱: yx3x@mail.ustc.edu.cn.

稿件信息: 2023 年 4 月 19 日来稿, 6 月 12 日接受, 稿件编辑: 王子达.

1 不可压欧拉方程在 $\mathbb{R}^d$ 中的 BKM 爆破准则

设 $d = 2, 3$, 给定速度场 $u = (u_1, \cdots, u_d) \in \mathbb{R}^d$, 我们定义其旋度为

$$\omega := \begin{cases} \nabla^\perp \cdot u = \partial_1 u_2 - \partial_2 u_1 & d = 2 \\ \nabla \times u = (\partial_2 u_3 - \partial_3 u_2, \partial_3 u_1 - \partial_1 u_3, \partial_1 u_2 - \partial_2 u_1) & d = 3 \end{cases}. \tag{1.1}$$

需要注意, 二维情况下的旋度是标量, 而三维情况下是向量.

下面叙述 Beale-Kato-Majda [5] 的主要结果.

定理 1.1 (BKM 爆破准则)**.** 设 $u(t,x) \in L^\infty([0,T); H^s(\Omega))$ 是不可压缩欧拉方程(0.1)的一个强解, 其中 $\Omega = \mathbb{R}^d$ 或 $\mathbb{T}^d$, $d = 2, 3$, $s > \frac{d}{2} + 1$, $T > 0$. 则

$$\limsup_{t \to T_-} \|u(t,\cdot)\|_{H^s} < \infty \text{ 当且仅当} \int_0^T \|\omega(t,\cdot)\|_{L^\infty} \,\mathrm{d}t < \infty. \tag{1.2}$$

换言之, 全空间欧拉方程(0.1)的解于 T 时刻在 Sobolev 空间 H^s 中爆破, 当且仅当其旋度在此时间段的时间积分产生爆破. 反之, 若旋度的时间积分在 T 时刻仍然有限, 则欧拉方程(0.1)的解仍可以延续到 T 之后的某个时刻 $T + \eta$ $(\eta > 0)$.

在介绍 BKM 爆破准则的证明之前, 我们先说明为什么这个定理直接蕴含了二维全空间中欧拉方程解的整体适定性. 这是因为二维情况下, 速度场的旋度的 L^p 范数是守恒量. 实际上, 对(0.1)的第一个方程求 $\nabla^\perp \cdot$ 并代入旋度的定义(1.1)以及不可压缩条件 $\nabla \cdot u = 0$, 我们可以得到旋度的演化方程 $(\partial_t + u \cdot \nabla)\omega = 0$. 容易证明其 L^p 范数必是守恒量

$$\frac{\mathrm{d}}{\mathrm{d}t} \int \omega^p \,\mathrm{d}x = \int (\nabla \cdot u)\omega^p = 0, \quad \forall p \geq 1.$$

而二维欧拉方程的初值 $u_0 \in H^s$, $s > 2$, 据 Sobolev 嵌入定理知, 必有 $\|\omega\|_{L^\infty} = \|\nabla^\perp \cdot u_0\|_{L^\infty} < \infty$ 成立. 因此定理 1.1 中的判定量永远不会爆破, 从而二维欧拉方程的解可以一直延续下去. 同时, 简单的能量估计表明解的 H^s 范数至多只有 e^{e^t} 的增长速率.

定理 1.2 (2D 欧拉方程整体解)**.** 当维数 $d = 2$ 时, 方程(0.1)的解满足: 对任意 $t \geq 0$, $s > 2$, 有

$$\|u(t,\cdot)\|_{H^s} \leq (e + e\|u_0\|_{H^s})^{2\exp(Ct(1+\|u_0\|_{L^2}+\|\omega_0\|_{L^\infty}))}$$

其中 $C > 0$ 仅依赖于 s.

注 . 迄今为止, 人们并不知道全空间中欧拉方程解的双指数增长率 $O(e^{e^t})$ 是否真的能够达到. 在 2013 年末, Zlatoš [38] 显式地构造了一个初值, 其演化出的解具有单指数增长率 $O(e^t)$. 而对边值问题, 则有 Kiselev-Šverák 根据 "棋盘初值" 构造出的一个具体例子能做到双指数增长.

BKM 爆破准则的证明本质上是具有齐次核的 Calderón-Zygmund 奇异积分 (下称 C-Z 奇异积分) 的逐点估计. 在此, 我们只叙述大致的证明方法和所需的调和分析工具.

首先, 我们将 "爆破准则" 这一问题进行约化. 现固定 s, 令 $E_s(t) = \|u(t,\cdot)\|_{H^s}^2$, 要想导出关于 $E_s(t)$ 的爆破准则, 则需要从欧拉方程(0.1)自身出发导出一个 Grönwall 型不等式

$$E_s'(t) \leq K(t) E_s(t),$$

从而在 $E_s(0)$ 有限的情况下, $E_s(t)$ 在某一时刻爆破与否完全取决于 $K(t)$ 是否在该时刻爆破, 这是证明的第一步. 第二步则是去实现上面这个 Grönwall 型不等式, 这可以通过直接对欧拉方程求 ∂^s 并作能量估计, 所得结果为

$$E_s'(t) \le C\|\nabla u\|_{L^\infty} E_s(t). \tag{1.3}$$

第三步, 也就是最重要的一步, 我们需要把 $\|\nabla u\|_{L^\infty}$ 约化为旋度的对应范数 $\|\omega\|_{L^\infty}$, 这是因为旋度本身是一个**可观测量**, 并且 L^∞ 本身是**逐点范数**. 这个约化使得该爆破准则是**可观测的**.

那么, 如何将速度场 u(的导数) 从旋度 ω 中**还原出来**? 这个约化过程实际上就是物理中的毕奥-萨法尔定律 (Biot-Savart law): 给定一个向量场的旋度, 如何还原该向量场本身? 这可以从一个简单的向量恒等式做起, 以三维为例, 我们有

$$\nabla\times\omega = \nabla\times(\nabla\times u) = -\triangle u + \nabla(\underbrace{\nabla\cdot u}_{=0}) = -\triangle u,$$

因此我们可以把速度场写成

$$u = (-\triangle)^{-1}(\nabla\times\omega) \Leftrightarrow \hat{u}(\xi) = \frac{i\xi}{|\xi|^2}\times\hat{\omega}(\xi).$$

经过具体计算, 我们有

$$u(x) = \int_{\mathbb{R}^3} K_3(x-y)\times\omega(y)\,\mathrm{d}y,$$

其中 $K_3(x) = \frac{x}{4\pi|x|^3}$. 对二维情况, 我们类似地可以证明 $\hat{u}(\xi) = \frac{i\xi^\perp}{|\xi|^2}\hat{\omega}(\xi)$, 以及

$$u(x) = \int_{\mathbb{R}^2} K_2(x-y)\times\omega(y)\,\mathrm{d}y, \quad K_2(x) = \frac{x^\perp}{2\pi|x|^2}.$$

现在我们需要将 ∇u 约化为旋度 ω, 所以我们对上述两个卷积式求导数, 计算可得

- $d=2$:

$$\nabla u = \frac{\omega(x)}{2}\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix} + \frac{1}{2\pi}P.V.\int_{\mathbb{R}^2}\frac{1}{|x-y|^2}\sigma\left(\frac{x-y}{|x-y|}\right)\omega(y)\,\mathrm{d}y,$$

  其中 $\sigma(z) = \begin{bmatrix}2z_1z_2 & z_2^2-z_1^2\\ z_1^2-z_2^2 & -2z_1z_2\end{bmatrix}$ 满足 $\int_{S^1}\sigma(z)\,\mathrm{d}S_z = 1$. 从而上式右端是 C-Z 奇异积分算子.
- $d=3$: 对任意向量 $h\in\mathbb{R}^3$ 有

$$\nabla u(x)h = \frac{\omega(x)\times h}{3} - \frac{1}{4\pi}P.V.\int_{\mathbb{R}^3}\frac{1}{|x-y|^3}\left(\omega(y)\times h + 3\left(\frac{x-y}{|x-y|}\times\omega(y)\right)\frac{x-y}{|x-y|}\cdot h\right)\mathrm{d}y,$$

  其中第二项是向量值 C-Z 奇异积分算子, 因为有恒等式

$$3\int_{S^2}(z\times\omega)z\cdot h\,\mathrm{d}z = -\frac{4\pi}{3}\omega\times h = -\int_{S^2}\omega\times h\,\mathrm{d}z.$$

可以看见, 无论是二维还是三维, $|\nabla u(x)| \lesssim |\omega(x)| + |T\omega(x)|$ 总是成立的, 其中 T 是 C-Z 卷积型奇异积分算子. 稍具有调和分析基础知识的读者都应知道, 一般的 C-Z 奇异积分算子仅仅具有 $L^p\to L^p$ $(1<p<\infty)$ 的有界性. 但当积分核为度 $(-d)$ 齐次函数的时候, 我们有逐点估计.

设 $\Omega: S^{d-1}\to\mathbb{R}$ 是光滑函数并满足 $\int_{S^{d-1}}\Omega(y)\,\mathrm{d}S_y = 0$. 据此, 定义具有齐次核的 C-Z 奇异积分算子如下

$$Tf(x) = P.V.\int_{\mathbb{R}^d}\frac{\Omega(y/|y|)}{|y|^d}f(x-y)\,\mathrm{d}y. \tag{1.4}$$

引理 1.3. 设 $f \in L^2 \cap L^\infty \cap C^\alpha$, 其中 $\alpha \in (0,1)$. 则

$$\|Tf\|_{L^\infty} \lesssim_{d,\alpha,\Omega} \|f\|_{L^2} + \|f\|_{L^\infty}\left(1+\log^+\frac{[f]_{C^\alpha}}{\|f\|_{L^\infty}}\right). \tag{1.5}$$

进一步地, 对 $s > \frac{d}{2}$, 据 Sobolev 嵌入定理 $H^s \hookrightarrow C^\alpha$ $(\alpha \in (0,\min\{1, s-\frac{d}{2}\}))$, 我们可以将不等式改写为

$$\|Tf\|_{L^\infty} \lesssim_{d,s,\Omega} \|f\|_{L^2} + \|f\|_{L^\infty}\left(1+\log^+\frac{\|f\|_{H^s}}{\|f\|_{L^\infty}}\right). \tag{1.6}$$

再进一步, 若存在位势函数 $g \in L^2$ 使得 $f = \nabla g$, 则

$$\|Tf\|_{L^\infty} \lesssim_{d,s,\Omega} \|g\|_{L^2} + \|f\|_{L^\infty}\left(1+\log^+\frac{\|f\|_{H^s}}{\|f\|_{L^\infty}}\right). \tag{1.7}$$

现在我们可以证明 BKM 爆破准则了.

BKM 爆破准则的证明. 首先, 如果旋度爆破, 即 $\lim\limits_{t\to T_-}\int_0^t \|\omega(\tau,\cdot)\|_{L^\infty}\,\mathrm{d}\tau = +\infty$, 则根据 Sobolev 嵌入定理, 对任意 $s-1 > d/2$ 就必有 $\lim\limits_{t\to T_-}\int_0^t \|u(\tau,\cdot)\|_{H^s}\,\mathrm{d}\tau = +\infty$, 因此这一侧是简单的.

反过来, 我们要证明如果旋度在 T 时刻不爆破, 即 $\lim\limits_{t\to T_-}\int_0^t \|\omega(\tau,\cdot)\|_{L^\infty}\,\mathrm{d}\tau < +\infty$, 则解也不爆破, 即

$$\sup_{t\in[0,T]} \|u(t,\cdot)\|_{H^s} \le C(T) = C\left(\|u_0\|_{H^s}, T, \int_0^T \|\omega(\tau)\|_{L^\infty}\,\mathrm{d}\tau\right) < +\infty.$$

现在, 我们在引理 1.3 的式子(1.7)中取 $f=\omega$, 便有

$$\|T\omega\|_{L^\infty} \lesssim_{d,s} \|u\|_{L^2} + \|\omega\|_{L^\infty}\left(1+\log^+\frac{\|\omega\|_{H^s}}{\|\omega\|_{L^\infty}}\right).$$

由于 $|\nabla u(x)| \lesssim |\omega(x)| + |T\omega(x)|$ 总是成立的, 以及 u 的 L^2 范数是守恒量, 根据能量估计的结果 $E_s'(t) \le C\|\nabla u\|_{L^\infty}E_s(t)$ 我们有

$$E_s'(t) \le C\left(\|u_0\|_{L^2} + \|\omega\|_{L^\infty}\left(1+\log^+\frac{\|u\|_{H^s}}{\|\omega\|_{L^\infty}}\right)\right)E_s(t),$$

从而

$$E_s'(t) \le C(d,s)\left(1+\|u_0\|_{L^2} + \|\omega\|_{L^\infty}\left(1+\log^+\|u\|_{H^s}\right)\right)E_s(t).$$

利用常微分方程的积分因子法, 对微分不等式 $Y'(t) \le \{A + B(t)[1+\log(1+Y(t))]\}Y(t)$, 我们可以得到

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(Y(t)e^{-At}\right) \le B(t)\left(1+At+\log\left(1+Y(t)e^{-At}\right)\right)\left(Y(t)e^{-At}\right).$$

即

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\log\left(1+At+\log(1+Y(t)e^{-At})\right)\right) \le \frac{A}{1+At} + B(t).$$

两边积分并去掉对数之后可得

$$Y(t) \le e^{At}\left\{\exp\left[(1+At)\left((1+\log(1+Y(0)))\exp\left(\int_0^t B(\tau)\,\mathrm{d}\tau\right)-1\right)\right]-1\right\}.$$

现在取 $A = C(d,s)(1+\|u_0\|_{L^2})$ 以及 $B(t) = C(d,s)\|\omega(t)\|_{L^\infty}$ 即得我们想要的爆破准则(1.2). □

如上 BKM 爆破准则是对全空间的欧拉方程证明的. 但现实中流体经常是在一个带边区域内运动, 甚至需要考虑流体的自由边界演化问题 (例如水波), 因此我们需要推广该爆破准则到边值问题.

2 带边区域的情况

今假设 $\Omega \subset \mathbb{R}^3$ 是一个有界单连通区域, 并具有光滑边界 $\partial\Omega$. 欧拉方程在 Ω 中的初边值问题定义如下

$$\begin{cases} (\partial_t + u \cdot \nabla)u = -\nabla p & \text{in } [0,T] \times \Omega \\ \nabla \cdot u = 0 & \text{in } [0,T] \times \Omega \\ u \cdot N = 0 & \text{on } [0,T] \times \partial\Omega \\ u|_{t=0} = u_0(x) & \text{on } \{t=0\} \times \Omega. \end{cases} \tag{2.1}$$

其中第三个方程是欧拉方程固壁问题的标准边界条件, $u \cdot N = 0$ 表示流体与固壁 $\partial\Omega$ 没有法向的碰撞, 该条件也被称作滑动条件 (slip condition). 边值问题(2.1)的局部适定性由 Kato-Lai [19] 证明. 1993 年, Beale 的学生 Ferrari 证明了如下爆破准则 [12] .

定理 2.1. 设如上边值问题的初值 $u_0 \in H^s(\Omega)$, $s \geq 3$ 满足 $\nabla \cdot u_0 = 0$, $u_0 \cdot N|_{\partial\Omega} = 0$, $u \in C([0,T);H^s(\Omega))$ 为上述初边值问题的强解. 若 $t = T^*$ 是使得 $u \notin C([0,T^*];H^s(\Omega))$ 最小时间, 则在此时间有 $\int_0^{T^*} \|\omega(t,\cdot)\|_{L^\infty}\,\mathrm{d}t = +\infty$.

上述结论看上去和 BKM 爆破准则毫无差异, 然而证明过程却有很大的不同, 尤其是如下两个方面

- 边界 $\partial\Omega \neq \varnothing$: 需要椭圆估计来控制压力 p. 注意, 压力 p 在 Ω 内满足椭圆方程 $-\triangle p = \partial_i u_j \partial_j u_i$, 但在边界上并没有 p 的边界条件. 因此我们需要将(2.1)限制在边界上并投影到法向来硬造一个 Neumann 边界条件. 而在 $\Omega = \mathbb{R}^d$ 时, 我们可以对 $\triangle$ 求逆 (利用傅立叶刻画) 直接用 $\|\nabla u\|_{s-2}$ 控制 $\|p\|_{H^s}$.
- 从 ∇u 到旋度 ω 的约化: 在 $\Omega = \mathbb{R}^d$ 时, 我们利用 Calderón-Zygmund 奇异积分理论完成这个约化过程, 但现在区域并非全空间, 奇异积分算子甚至无法定义. 此时需要用散度-旋度方程组来进行椭圆估计.

上述第一个问题并不困难, 直接计算可得

$$-\triangle p = \partial_i u_j \partial_j u_i \text{ in } \Omega, \quad \left.\frac{\partial p}{\partial N}\right|_{\partial\Omega} = -u_i \partial_i u_j N_j.$$

对有界区域内带 Neumann 边界条件的椭圆方程 $-\triangle U = f$ in Ω, $\frac{\partial U}{\partial N}|_{\partial\Omega} = g$, 我们有

$$\forall s \geq 2,\ \|U\|_{H^s(\Omega)} \lesssim \|f\|_{H^{s-2}(\Omega)} + |g|_{H^{s-1.5}(\partial\Omega)}.$$

利用该估计、边界条件 $u \cdot N|_{\partial\Omega} = 0$、以及迹定理, 我们仍然可以得到 $\|\nabla p\|_{H^s} \leq C\|u\|_{H^s}\|u\|_{W^{1,\infty}}$.

解决了这个问题之后, 对固壁问题(2.1), 我们仍可以导出能量估计 $E_s'(t) \leq C\|u\|_{W^{1,\infty}} E_s(t)$. 于是, 我们只要解决上述第二个问题, 也就是从 ∇u 到旋度 ω 的约化过程.

上面提到, 我们的约化过程需要的是对散度-旋度方程组的椭圆估计. 具体而言, 对向量场 u, 我们已有

$$\nabla \times u = \omega, \quad \nabla \cdot u = 0 \text{ in } \Omega, \quad u \cdot N|_{\partial\Omega} = 0.$$

为此, 我们考虑一个更一般的椭圆方程组

$$\begin{cases} \nabla \times u - \nabla\varphi = \zeta & \text{in } \Omega \\ \nabla \cdot u = \sigma & \text{in } \Omega \\ u \cdot N = \beta & \text{on } \partial\Omega \\ \varphi = \gamma & \text{on } \partial\Omega. \end{cases} \tag{2.2}$$

首先, 为什么说这是椭圆方程组? 实际上对第一个方程求旋度, 并利用向量恒等式 $\nabla\times(\nabla\times u) = -\Delta u+\nabla(\nabla\cdot u)$ 可得 $-\nabla u = -\nabla\sigma + \zeta$. 另一方面, 对第一个方程求散度, 则有 $-\Delta\varphi = \nabla\cdot\zeta$. 所以无论是向量函数 u 还是标量函数 φ 都由它们满足的椭圆方程以及对应的边界条件所确定.

现在, 令 $v = (u,\varphi)$, $\phi = (\zeta,\sigma)$, $\chi = (\beta,\gamma)$, 则上述方程组(2.2)可以写成

$$\mathcal{L}v = \phi \text{ in } \Omega, \quad \mathcal{B}v|_{\partial\Omega} = \chi,$$

其中 $\mathcal{L}$ 是一个一致椭圆算子, B 是对应的边界矩阵. Ferrari 证明, 对这样的椭圆系统, 我们仍然可以用牛顿位势来表示方程的解 (这实际上就是奇异积分的动机, 即研究椭圆方程的 $W^{2,p}$ 估计), 并成立如下结论.

引理 2.2. 假设 $\Omega\subset\mathbb{R}^3$ 是一个有界单连通区域, 并满足 $\partial\Omega\in C^{2,\gamma}$ $(0<\gamma<1)$, 若 $v\in H^3$ 是(2.2)在 $\phi\in H^2(\Omega)$ 情况下的解并且 $\chi=0$, 则有估计

$$\|v\|_{W^{1,\infty}(\Omega)} \le C\left(1+\log^+\frac{\|\phi\|_{H^2(\Omega)}}{\|\phi\|_{L^\infty(\Omega)}}\right)\|\phi\|_{L^\infty(\Omega)}.$$

特别地, 令 $\varphi=0$, $\zeta=\omega$, $\sigma=0$, $\beta=\gamma=0$, 我们就可以证明想要的约化

$$\|u(t,\cdot)\|_{W^{1,\infty}(\Omega)} \le C\left[\left(1+\log^+\|u(t,\cdot)\|_{H^s(\Omega)}\right)\|\omega\|_{L^\infty(\Omega)}+1\right].$$

以上是对 Ferrari [12] 证明的简要概述. 现在, 让我们考虑一个更深刻的问题: 欧拉方程自由边界问题是否也有类似的爆破准则? 实际上, 这个问题更加具有物理意义: 可以想象现实生活中的海洋, 其表面总是在运动的而非静止的. 自然地, 我们问: 欧拉方程的自由边界问题是否具有小初值整体解? 若无, 是否能导出一个爆破准则? 关于第一个问题, 过去二十余年内已经有了长足的发展. 而第二个问题, 至今并没有一个完美的答案.

3 自由边界问题的进展与困难

我们考虑两个情况: 重力水波、理想液滴. 前者对应的区域是 $\Omega(t) := \mathbb{R}^{d-1}\times(-b,\psi(t))$ (其中 b 可以是 ∞, 即代表无穷深), 后者对应的区域是 $\mathcal{D}(t) := \mathbb{T}^{d-1}\times(-b,\psi(t))$. 其中流体的自由界面分别为 $\{x\in\mathbb{R}^d : x_d = \psi(t)\}$ 和 $\{x\in\mathbb{T}^d : x_d=\psi(t)\}$, 为了记号简便, 我们均将其记为 $\Sigma(t)$. 而流体的固定底边则记作 Σ_b. 现在我们引入不可压缩欧拉方程的自由边界问题

$$\text{(重力水波方程)}\quad \begin{cases} (\partial_t + u\cdot\nabla)u = -\nabla p - ge_d & \text{in } [0,T]\times\Omega(t) \\ \nabla\cdot u = 0 & \text{in } [0,T]\times\Omega(t) \\ u\cdot e_d = 0 & \text{on } [0,T]\times\Sigma_b \\ \psi_t = u\cdot\mathbf{N} & \text{on } [0,T]\times\Sigma(t) \\ p = -\sigma\overline{\nabla}\cdot\left(\dfrac{\overline{\nabla}\psi}{\sqrt{1+\left|\overline{\nabla}\psi\right|^2}}\right) & \text{on } [0,T]\times\Sigma(t). \end{cases} \tag{3.1}$$

$$
(\text{理想液滴}) \quad \begin{cases} (\partial_t + u \cdot \nabla) u = -\nabla p & \text{in } [0,T] \times \mathcal{D}(t) \\ \nabla \cdot u = 0 & \text{in } [0,T] \times \mathcal{D}(t) \\ u \cdot e_d = 0 & \text{on } [0,T] \times \Sigma_b \\ \psi_t = u \cdot \mathbf{N} & \text{on } [0,T] \times \Sigma(t) \\ p = -\sigma \overline{\nabla} \cdot \left(\dfrac{\overline{\nabla}\psi}{\sqrt{1 + \left|\overline{\nabla}\psi\right|^2}} \right) & \text{on } [0,T] \times \Sigma(t). \end{cases} \tag{3.2}
$$

这其中 ge_d 代表竖直方向的重力, $\mathbf{N} := (-\overline{\partial}_1\psi, \cdots, -\overline{\partial}_{d-1}\psi, 1)^\top$ 是自由界面的法向量 (并非单位法向量, 但用这种法向量更方便计算), 常数 $\sigma \geq 0$ 称作表面张力系数. 特别地, 当 $\sigma = 0$ 时, 我们需要 Rayleigh-Taylor 条件

$$
-\mathbf{N} \cdot \nabla p|_{\Sigma(t)} \geq c_0 > 0 \tag{3.3}
$$

来保证局部适定性. 去掉该条件, 在 $\sigma = 0$ 时, 欧拉方程的自由边界问题不是适定的, 参见 Ebin [10] 构造的反例.

不可压缩欧拉方程自由边界问题的局部适定性并不是一件非常显见的事情. 与固壁问题不同, 自由边界问题的一个显著的困难是**自由界面的正则性不高, 并且依赖于速度场.** 例如 $u \in H^s(\Omega(t))$, 那么自由界面 ψ 在 $\sigma = 0$ 时至多只有 H^s 正则性, 在 $\sigma > 0$ 时至多只有 $H^{s+1.5}$ 正则性.

局部适定性的证明首先由华人数学家邬似珏在 1997、1999 年用调和分析、复分析、Clifford 分析的方法做出无旋流的情况 [33, 34]. 对有旋且 $\sigma = 0$ 情况, Christoudoulou-Lindblad [6] 做出了先验估计, Lindblad [22] 利用 Nash-Moser 迭代做出了局部适定性, 而不用 Nash-Moser 迭代的证明则首先由 Coutand-Shkoller [8] 给出. 当 $\sigma > 0$ 时, Iguchi [16] 首先证明了无旋情况的局部适定性, 有旋流的适定性则由 Coutand-Shkoller [8] 给出. 此外, 还有非常多关于局部适定性的文章 [4, 37, 29, 1, 31, 32] 等, 在此不一一列举.

对自由边界, 我们基本只能考虑半空间区域水波方程的不可压缩无旋流的小初值整体解, 即方程(3.1). 这样考虑是出于三个原因. 第一, 我们对初始界面的扰动必须要小, 否则可能产生界面自交、界面崩塌、喷出单独水滴等奇异性, 例如 Castro, Córdoba, Fefferman, Gancedo, Gómez-Serrano [7] 关于界面自交奇异性的刻画. 第二, 我们考虑无旋流是因为旋度自身仅仅满足传输方程, 并没有关于时间的衰减, 因此想证明有旋流的整体解几乎是不可能. 这里要注意, 无旋条件仅仅是对欧拉方程初值的假设, 而非一个额外的演化方程, 我们可以证明只要欧拉方程初值无旋则解一定无旋, 但这个性质对更复杂的流体方程并不一定成立. 第三, 我们考虑半空间模型 $\Omega(t) := \mathbb{R}^{d-1} \times (-b, \psi(t))$ 是因为其自由界面是 $\mathbb{R}^{d-1}$, 因此可以进行色散估计 (衰减估计). 在有界区域上, 色散估计未必成立.

无旋流整体解方面的首个突破仍然是华人数学家邬似珏在 2009、2011 年证明的结果 [35, 36], 此后人们利用更加精细的调和分析方法, 尤其是仿积 (paraproduct) 理论, 建立了更优的结果或者更简单的证明, 参见 [13, 2, 17, 15, 9] 等文章. 但迄今为止, 人们并不知道当 $d = 2$, $\sigma > 0$ 时, 重力水波方程(3.1)是否有小初值整体解. 而当区域为 $\mathcal{D}(t) := \mathbb{T}^{d-1} \times (-b, \psi(t))$ 时, 囿于水平方向是有界区域, 迄今为止人们也不知道能不能做出小初值长时间解.

因此, 寻求一个爆破准则是一个更可行的想法. 由于在全空间内和在固壁区域内, 我们已有 BKM 爆破准则 [5, 12], 所以我们希望自由边界问题的爆破准则满足如下三个要求:

1. 区域固定时, **能"退化"到固壁问题**的 BKM 爆破准则 [12].
2. 判定方程解爆破与否的量**必须是可观测量**.

3. 更进一步地, 能否改进 [12] 中的正则性 H^3 到局部适定性的**最低正则性** $H^{\frac{d}{2}+1+\varepsilon}(\Omega)$? 以及自由边界的**最低正则性** $\psi \in C^{1,\gamma}$? (注: 该自由界面正则性是数值模拟发现的, 低于该正则性, 流体自由界面会产生奇异性)

遗憾的是, 这三个问题至今并没有一个完美的答案, 人们也许需要一些全新的想法来同时解决上述三个问题. 笔者在此列举一些已有的结果, 并说明它们的不足之处, 以及提出笔者本人认为可能可行的改进方法.

或许是出于方便计算的考量, 以往所有的工作均考虑的是理想液滴的运动 (而非重力水波), 并且绝大多数均是考虑 $\sigma = 0$ 的情况. 即不带表面张力的自由边界问题(3.2), 而非重力水波方程(3.1). 在此, 我们需要注意, 当 $\sigma = 0$ 时, 为了能让解一直延续下去, 我们需要 Rayleigh-Taylor 符号条件(3.3)一直成立到解的爆破时刻之前, 即

$$\inf_{t\in[0,T^*]} -\mathbf{N}\cdot\nabla p|_{\Sigma(t)} \geq c_1 > 0 \tag{3.4}$$

其次, 流体的自由边界不可以触底, 我们需要

$$\inf_{t\in[0,T^*]} \psi(t) > -b. \tag{3.5}$$

首个显式的结果是王超-章志飞-赵威任-郑云瑞 [31] 在证明 $\sigma = 0$ 局部适定性时顺带得到的爆破准则, 其雏形实际上已经在 Christoudoulou-Lindblad [6] 中出现. [31] 证明了解在如下两个条件之一失效的时候会产生爆破

$$\sup_{t\in[0,T^*]} \left(|\mathcal{H}(t)|_{L^p\cap L^2(\Sigma(t))} + \|u(t)\|_{W^{1,\infty}(\mathcal{D}(t))}\right) < +\infty, \quad p \geq 6, \tag{3.6}$$

$$\inf_{t\in[0,T^*]} -\mathbf{N}\cdot\nabla p|_{\Sigma(t)} \geq c_1 > 0. \tag{3.7}$$

可见, 这仅仅是一个由适定性结果得到的爆破准则, 没有约化到旋度, $L^p \cap L^2$ 范数也并非可观测的量 (现实中只能观测逐点行为, 不可能观测积分. 而且即使某一点爆破, 积分也未必爆破), 最低正则性就差得更远了. 因此, 这个爆破准则未解决如上三个问题中的任何一个. 然而这在当时 (2014 年底) 已经是最佳结果了, 可见这个问题之困难. 一个更简单的叙述可以看王超-章志飞写的综述 [30].

第二个结果由 Dan Ginsberg [14] 给出 (2018 年底), 他给出的判定量是

$$\int_0^T \left(\|\omega(t)\|^2_{L^\infty(\mathcal{D}(t))} + |\nabla u(t)|_{L^\infty(\Sigma(t))} + \|\mathcal{N}(u|_{\Sigma(t)})\|_{L^\infty(\Sigma(t))} + |\nabla_N D_t p(t)|_{L^\infty(\Sigma(t))}\right) \mathrm{d}t, \tag{3.8}$$

以及若干边界项和(3.4). 其中 $\mathcal{N}$ 是区域 $\Omega(t)$ 的 Dirichlet-to-Neumann 算子, $D_t := \partial_t + u\cdot\nabla$ 为物质导数. 首先, 这里的判定量是 $\|\omega(t)\|^2_{L^\infty(\mathcal{D}_t)}$ 而不是线性, 这会让人质疑该结果的正确性. 因为如果能量估计里面 $\|\nabla u(t)\|_{L^\infty(\mathcal{D}_t)}$ 不是线性的, 那么在进行旋度约化后得到的 Grönwall 型微分不等式的时候旋度项 $\|\omega(t)\|_{L^\infty(\mathcal{D}_t)}$ 也不是线性的, 进而解这个微分不等式的时候得到的 $\|\omega(t)\|_{L^\infty(\mathcal{D}_t)}$ 的幂次并不一定和能量估计中的幂次相同. 第二, 假设该结果正确, 由于平方项的存在, 其无论如何不可能约化为 Ferrari 证明的固壁问题的 BKM 爆破准则 [12]. 第三, $|\nabla_N D_t p(t)|_{L^\infty(\Sigma(t))}$ 并不是一个可观测的量, 也不会在固壁问题中消失. 另外, 这篇文章用的是拉格朗日坐标系来固定自由区域, 这实际上会在能量估计中引进大量不好控制的非线性项. 所以说, [14] 除了进行了旋度约化之外, 离我们所需的目标实则更加遥远, 甚至不如 [31] 的结果.

第三个结果是 Julin-La Manna [18] 关于 3D 带电液滴的先验估计以及爆破准则的证明 (2021 年底). 这篇文章的作者为了将界面正则性降到最佳, 即 $\psi \in C^{1,\gamma}$, 直接将该方程放在欧拉坐标系里面研究, 即并不将自由边界问题约化为固定边界问题. 但是这样做的代价是他们没有进行旋度的约化. 更致命的是, 他们并没有用这个能量证明解的局部适定性, 等于说连这样的解存在不存在都不知道.

这里需要注意, 对于理想流体的自由边界问题, 即使做出先验估计也不能直接得出局部适定性, 迭代求解非线性问题并非易事. 实际上, 利用欧拉坐标系也许有利于计算低正则性估计, 但求解是相当困难的: 在迭代的过程中, 除了需要迭代变量, 还需要迭代流体区域! 这是因为不同的速度场、不同的线性化方程必定会演化出不同的流体区域! 若不通过固定住边界区域, 将 "区域的运动" 转化为具体可计算的量 (例如拉直映射、拉格朗日坐标系中速度场的流等等), 那么在迭代求解过程中就必须去估计**相邻两组逼近解对应区域的差集中的能量**, 其难度是无法想象的.

笔者的上述评论并非批评这三个结果不好. 实际上, 从问题的提出, 对困难点的分析, 就可以看出证明自由边界问题爆破准则之困难. 除此之外, 证明爆破准则看似是基于局部适定性的能量估计做进一步的约化, 但对自由边界问题而言, 从局部适定性到爆破准则存在一个较大的跨越: 证明爆破准则必须得到关于 $E_s(t)$ 线性的能量估计, 即 $E_s'(t) \le K(t)E_s(t)$ 而非证明局部适定性仅需的非线性估计 $E_s'(t) \le K(t)P(E_s(t))$, 否则我们解不了非线性的 Grönwall 型微分不等式; 而自由界面的出现恰恰会导致大量带 ψ 的非线性项出现在能量估计中!

在介绍上述三个结果之前, 笔者提到了一句

> "或许是出于方便计算的考量, 以往所有的工作均考虑的是理想液滴的运动 (而非重力水波), 并且绝大多数均是考虑 $\sigma = 0$ 的情况. 即不带表面张力的自由边界问题(3.2), 而非重力水波方程(3.1)."

与上述工作不同, 笔者认为**考虑 $\sigma = 0$ 的简单情况反而不可能得到一个 "完美的" 爆破准则.** 这个看似反常的想法实际上基于一个非常简单的观察. 相较固壁问题而言, 自由边界问题在 $\sigma = 0$ 时需要保持 Rayleigh-Taylor 符号条件(3.3)一直成立! 需要注意的是, 这个条件仅仅是对初值的限制, 而不是方程(3.1),(3.2)自带的边界条件. 在证明局部适定性的时候, 我们可以用 $D_t p$ 的界来证明这个条件在短时间内成立. 然而, 要使解一直延续下去, 我们就必须考虑 Rayleigh-Taylor 符号本身的演化, 这也解释了为什么 [31, 14] 的结果中出现了 $D_t p$ 这种项, 为什么 [14] 中会出现旋度的非线性项. 实际上, 对压力 p 满足的椭圆方程求物质导数, 得到的内部方程 $\Delta D_t p = \cdots$ 中的右端项是一个关于 ∇u 的三次项而不是二次项, 从而我们不可能得到关于 ∇u 线性的能量估计. 一旦能量估计丢失线性, 我们就不可能得到一个 "能退化为固壁问题的" 爆破准则!

因此, 要想得到一个完美的爆破准则, 应当考虑 $\sigma > 0$ 的情况而非 $\sigma = 0$, 从而避开研究 Rayleigh-Taylor 符号的演化. 笔者的合作者罗辰昀、周凯近期 (2023 年 2 月) 就证明了一个这样的爆破准则 [26]: 对(3.2)在 $d = 3$, $\sigma > 0$ 的情况, $C([0,T*); H^3(\Omega))$ 中的解 (u,ψ) 在 T^* 时刻爆破, 当且仅当下面三个量中的任一个量爆破:

a. (自由界面的逐点范数爆破)

$$\lim_{t\nearrow T^*} K(t) = +\infty,\ \ K(t) := |\psi(t)|_{C^3} + |\psi_t(t)|_{C^3} + |\psi_{tt}(t)|_{H^{1.5}} + |\bar{u}(t)|_{W^{1,\infty}}, \tag{3.9}$$

b. (旋度爆破)

$$\int_0^{T^*} \|\omega(t)\|_{L^\infty}\,\mathrm{d}t = +\infty, \tag{3.10}$$

c. (自由界面触底或不再是图像)

$$\lim_{t\nearrow T^*}\left(\frac{1}{\partial_3\varphi(t)} + \frac{1}{b - |\psi(t)|_\infty}\right) = +\infty. \tag{3.11}$$

其中 $\varphi(t,x) = x_3 + \chi(x_3)\psi(t,x_1,x_2)$ 是 ψ 在内部的延拓, φ 也是拉直边界所需的微分同胚.

可以看见, 上述结果至少做到了我们想要的前两个要求: 可退化到固壁问题的结果、以及所有的量都是和观测的. 其中需要提到的是, 固壁问题即假设 $\psi = 0$, 这样的话 $K(t) = 0$, 其中 $|\bar{u}(t)|_{W^{1,\infty}}$ 出现也是因为自由边界 $\psi_t \neq 0$, 在固壁问题中并不会有这样的项. 但是, [26] 并没有解决我们想要的第三点, 即把正则性降到最低.

从 $K(t)$ 的定义来看, 这个爆破准则反而出现了一些更高阶的项, 比如边界的 C^3 范数. 之所以出现这样的高阶项, 是因为在估计压力 p 的时候采用了 Dirichlet 边界条件而非 Neumann 边界条件. 这样做的好处是不需要引进速度场的时间导数估计, 但坏处是平均曲率项自身带有二阶导数导致能量的阶数太高. 幸运的是, 利用表面张力条件和 Schauder 估计 (参考 [28]), 我们可以将 $K(t)$ 中的量用存在性能量控制住. 因此, 这个爆破准则的判定量 $K(t)$ 与局部适定性的空间是匹配的.

笔者与合作者们在讨论之后均认为, 如果想要得到最佳正则性的爆破准则, 仍然需要采用 [26] 的方程写法而不是笔者之前与合作者证明可压重力水波适定性 [24] 时所用的拉格朗日坐标系. 实际上, [26] 对方程本身的约化采用的是笔者与罗辰昀研究可压毛细-重力水波 [25] 一文中的方法, 而该方法是推广了 Masmoudi-Roussét [27] 研究 Navier-Stokes 方程自由边界问题的无粘极限时采用的 Alinhac good unknown 方法 [3]. 但是, Neumann 边界条件的椭圆估计也许还是必须的, 否则高阶量 $K(t)$ 无法避开. 除此之外, 在低正则性的框架下, 大量的 "低阶余项" 不再变得低阶, 需要用 Kato-Ponce 型乘积估计 [20] 来解决, 这在笔者之前的工作中已经出现过 [23]. 但是, 如何合适地结合这些技术, 仍旧是非常困难的问题.

除此之外, 笔者与合作者认为, 更有意义的问题是考虑二维重力水波方程的爆破准则而非有界区域的情况. 一方面, 区域的有界性限制了色散估计, 从而不能期待有界区域问题(3.2)的长时间解存在性. 另一方面, 二维重力水波在 $\sigma > 0$ 时仍然不知道是否有小初值整体解. 若能做出二维重力水波在 $\sigma > 0$ 时的爆破准则, 结合二维情况下旋度守恒 (从而不可能爆破) 的性质, 便可以断言

"二维重力-毛细水波方程的解发生爆破, 只可能是边界能量首先爆破, 不可能是内部先爆破! "

(此处毛细是指 $\sigma > 0$) 这便展示出二维和三维情况的一个显著区别. 进一步地, 这样的爆破准则一旦成立, 人们可以更倾向猜测二维毛细-重力水波的小初值整体解仍是有可能证明的, 毕竟初始时刻对边界和速度场的小扰动, 没有理由在有限时刻增长到无穷大, 因此爆破准则中的判定量应当是一直有限的.

然而, 水波问题(3.1)是在无界区域中研究. 相较于有界区域, 无界区域的各类椭圆估计更加困难, 尤其是 Schauder 型估计. 即便假设初值是局部化的 (即几乎集中在一个紧集内), 我们也无法直接截断流体区域. 如何克服无界区域的困难, 重力项是否会产生干扰, 都是悬而未决的问题. 笔者期待在将来能对该问题做出本质突破.

参考文献

[1] Alazard, T., Burq, N., Zuily, C. On the Cauchy problem for gravity water waves. *Invent. Math.*, **198**(1), 71–163. 2014.

[2] Alazard, T., Delort, J.-M. Global solutions and asymptotic behavior for two-dimensional gravity water waves. *Ann. Sci. Éc. Norm. Supér.,* **48**(5), 1149–1238. (2015)

[3] Alinhac. S. Existence d'ondes de raréfaction pour des systèmes quasi-linéaires hyperboliques multidimensionnels. *Commun. Partial Differ. Equ.,* **14**(2), 173–-230.(1989)

[4] Ambrose, D., Masmoudi, N. The zero surface tension limit of two-dimensional water waves. *Commun. Pure Appl. Math.*, 58(10), 1287-1315. (2005)

[5] Beale, J. T., Kato, T., A. Majda. Remarks on the breakdown of smooth solutions for the 3-D Euler equations. *Commun. Math. Phys.*, **94**(1), 61-66. (1984)

[6] Christodoulou, D., Lindblad, H. On the motion of the free surface of a liquid. *Commun. Pure Appl. Math.*, **53**(12), 1536–1602, 2000.

[7] Castro, A., Córdoba, D., Fefferman, C., Gancedo, F., Gómez-Serrano, J. Interface evolution: water waves in 2-D. *Ann. Math.*, **178**(3), 1061-1134. (2013)

[8] Coutand, D., Shkoller, S. Well-posedness of the free-surface incompressible Euler equations with or without surface tension. *J. Amer. Math. Soc.*, **20**(3), 829–930. (2007)

[9] Deng, Y., Ionescu, A. D., Pausader, B., Pusateri, F. Global solutions of the gravity-capillary water-wave system in three dimensions. *Acta Math.*, **219**(2), 213–402. (2017)

[10] Ebin, D. G. The equations of motion of a perfect fluid with free boundary are not well-posed. *Commun. Partial Differ. Equ.*, **12**(10), 1175–1201. (1987)

[11] Elgindi, T. Finite-time singularity formation for $C^{1,\alpha}$ solutions to the incompressible Euler equations on $\mathbb{R}^3$. *Ann. Math.*, **194**(3), 647-727. (2021)

[12] Ferrari. A. B. On the blow-up of solutions of the 3-D Euler equations in a bounded domain. (English summary) *Comm. Math. Phys.* **155**(2), 277–294. (1993)

[13] Germain P., Masmoudi, N., Shatah J. Global solutions for the gravity water waves equation in dimension 3. *Ann. Math.*, **175**(2), 691–754. (2012)

[14] Ginsberg. D. On the breakdown of solutions to the incompressible Euler equations with free surface boundary. *SIAM J. Math. Anal.* **53**(3), 3366–3384. (2021)

[15] Ifrim. M, Tataru, D. Two-dimensional water waves in holomorphic coordinates II: global solutions. *Bulletin de la Société mathématique de France*, **144**(2), 366–394. (2016)

[16] Iguchi. T. Well-posedness of the initial value problem for capillary-gravity waves. *Funkcial. Ekvac.*, 44(2):219–242. (200)1

[17] Ionescu, A. D., Pusateri, F. Global solutions for the gravity water waves system in 2D. *Invent. Math.* **199**, 653–804. (2015)

[18] Julin, V., La Manna, D. A. A priori estimates for the motion of charged liquid drop: A dynamic approach via free boundary Euler equations. *arXiv:2111.10158v2*, preprint (2021).

[19] Kato, T., Lai, C. Y. Nonlinear Evolution Equations and the Euler Flow. *J. Funct. Anal.*, **56**, 15-28. (1984)

[20] Kato, T., Ponce, G. Commutator estimates and the Euler and Navier-Stokes equations. *Commun. Pure Appl. Math.*, **41**(7), 891–907. (1988)

[21] Kiselev, A., Šverák, V. Small scale creation for solutions of the incompressible two dimensional Euler equation. *Ann. Math.*, 180(3), 1205-1220. (2014)

[22] Lindblad, H. Well-posedness for the motion of an incompressible liquid with free surface boundary. *Ann. Math.*, **162**(1), 109-194. (2005)

[23] Luo, C., Zhang, J. A Regularity Result for the Incompressible Magnetohydrodynamics Equations with Free Surface Boundary. *Nonlinearity*, **33**(4), 1499-1527. (2020)

[24] Luo, C., Zhang, J. Local Well-posedness for the Motion of a Compressible Gravity Water Wave with Vorticity. *J. Differ. Equ.*, **332**, 333-403. (2022)

[25] Luo, C., Zhang, J. Compressible Gravity-Capillary Water Waves with Vorticity: Local Well-Posedness, Incompressible and Zero-Surface-Tension Limits. *arXiv:2211.03600*, preprint. (2022)

[26] Luo, C., Zhou, K. A Generalized Beale-Kato-Majda Breakdown Criterion for the free-boundary problem in Euler Equations with Surface Tension. *arXiv:2302.08125*, preprint. (2023)

[27] Masmoudi, N., Rousset, F. Uniform Regularity and Vanishing Viscosity Limit for the Free Surface Navier-Stokes Equations. *Arch. Rational Mech. Anal.*, **223**(1), 301-417. (2017)

[28] Nardi. G. Schauder estimate for solutions of Poisson's equation with Neumann boundary condition. *L'Enseignement Mathématique*, **60**(3), 421–435. (2015)

[29] Shatah, J., Zeng, C. Local well-posedness for fluid interface problems. *Arch. Rational Mech. Anal.,* **199**(2), 653-705. (2011)

[30] Wang, C., Zhang, Z. Breakdown criterion for the water-wave equation. *Sci. China Math.,* **60**, 21–58. (2017)

[31] Wang, C., Zhang, Z., Zhao, W., Zheng, Y. Local well-posedness and break-down criterion of the incompressible Euler equations with free boundary. *Mem. Am. Math. Soc.* **270**, no. 1318, 119 pp. (2021)

[32] Wang, Y., Xin, Z. Vanishing Viscosity and Surface Tension Limits of Incompressible Viscous Surface Waves. *SIAM J. Math. Anal.*, **53**(1), 574-648. (2021)

[33] Wu, S. Well-posedness in Sobolev spaces of the full water wave problem in 2-D. *Invent. Math.*, **130**(1), 39-72. (1997)

[34] Wu, S. Well-posedness in Sobolev spaces of the full water wave problem in 3-D. *J. Am. Math. Soc.* **12**(2), 445–495. (1999)

[35] Wu, S. Almost global wellposedness of the 2-D full water wave problem. *Invent. Math.* **177**(1), 45–135. (2009)

[36] Wu, S. Global wellposedness of the 3-D full water wave problem. *Invent. Math.* **184**(1), 125–220. (2011)

[37] Zhang, P., Zhang, Z. On the free boundary problem of three-dimensional incompressible Euler equations. *Commun. Pure Appl. Math.*, **61**(7), 877–940. (2008)

[38] Zlatoš, A. Exponential growth of the vorticity gradient for the Euler equation on the torus. *Adv. Math.* **268**, 396-403. (2015)

The Control of Geometric Quantities Under Scalar Curvature Conditions and Rigidity Phenomena

Li Xuanyu

Abstract

In this note, we survey several sharp inequalities involving the scalar curvature. In particular, we characterize some rigidity cases given the condition of scalar curvature and the existence of certain kinds of minimal surfaces.

1 Scalar Curvature and Minimal Surfaces

1.1 Scalar Curvature

Given a Riemannian manifold (M^n, g), its scalar curvature R_g is defined as the trace of the Ricci tensor. Unlike the sectional and Ricci curvatures, there is no such an direct interpretation of the scalar curvature on the size of manifolds such as the diameter and volume. However, it still have some local interpretations. For example, we have the volume expansion:

$$\mathrm{Vol}(B_r(p), g) = \omega_n r^n - \omega_n \frac{R_g(p)}{6(n-2)} r^{n+2} + O(r^{n+4}), \forall p \in M$$

where $B_r(p)$ is the geodesic ball centered at p in $M$and ω_n is the volume of the unit ball in $\mathbb{R}^n$. As a result, for two manifolds (M, g) and (M', g') and two points $p \in M, p' \in M'$, we have

$$R_g(p) > R_{g'}(p') \iff \mathrm{Vol}(B_r(p)) < \mathrm{Vol}(B_r(p')) \ \text{ for } \ r > 0 \ \text{ sufficiently small.}$$

There is a similar expansion of the total mean curvature of the geodesic sphere, see [14].

At the end of this subsection we fix some notations.

Definition 1.1. *Given a Riemannian manifold* (M, g)*.*

(1) The Levi-Civita connection of M *will be denoted by* D*.*

(2) The Gauss, sectional, Ricci and scalar curvature will be denoted by $K, \mathrm{Sec}, \mathrm{Ric}, R$ *respectively. To avoid confusion, we will sometimes use the subscript to denote the dependence on the metric and superscript for the manifold. For example, the scalar curvature of* Σ *in* (1) *will be denoted by* R_g^Σ*.*

1.2 Minimal Surfaces

Next we recall some basic facts from minimal surfaces. Given an embedded compact hypersurface $\Sigma^n \subset (M^{n+1}, g)$, its **second fundamental form** is a symmetric bi-linear map on $T_p\Sigma$ which takes value in

作者信息: 李轩宇, 2019 级数学科学学院, 邮箱: lixuanyu@mail.ustc.edu.cn.
稿件信息: 2022 年 12 月 17 日来稿, 2023 年 5 月 15 日修改, 5 月 17 日接受, 稿件编辑: 刘畅.

$T_p^\perp\Sigma$,

$$\mathrm{I\!I}: T_p\Sigma \times T_p\Sigma \to T_p^\perp\Sigma, \mathrm{I\!I}(X,Y) := -(D_X Y)^\perp.$$

The **mean curvature vector** H is defined as the trace of $\mathrm{I\!I}$. In most cases, we assume there is a global unit normal ν to Σ and, with a little abuse of notation, denote also by $\mathrm{I\!I}$ the symmetric 2-tensor

$$\mathrm{I\!I}(X,Y) = -\langle D_X Y, \nu\rangle = \langle D_X \nu, Y\rangle$$

and so as H. Now, if Σ is two-sided, for $f \in C^\infty(\Sigma)$, we denote by F_t the flow of $f\nu$, i.e. $\frac{\partial}{\partial t}F_t\big|_{t=0} = f\nu$. Set $\Sigma_t := F_t(\Sigma)$, the **first variation formula** is

$$\frac{d}{dt}\mathrm{Vol}(\Sigma_t)\bigg|_{t=0} = -\int_\Sigma fH d\sigma.$$

We say Σ is **minimal** if it is the critical point of the volume functional, which is equivalent to the fact that H vanishes identically. If Σ is minimal, then the **second variation formula** is given by

$$\frac{d^2}{dt^2}\mathrm{Vol}(\Sigma_t)\bigg|_{t=0} = \int_\Sigma |\nabla^\Sigma f|^2 - (\mathrm{Ric}(\nu,\nu) + |\mathrm{I\!I}|^2)f^2 d\sigma = -\int_\Sigma f(\Delta_\Sigma f + (\mathrm{Ric}(\nu,\nu) + |\mathrm{I\!I}|^2)f)d\sigma \tag{1.1}$$

where $\nabla^\Sigma, \Delta_\Sigma$ are the **gradient** and **Laplacian** of Σ respectively. Note that $\mathrm{Ric}(\nu,\nu)$ is independent of ν and hence is always smooth even when Σ is one-sided. It follows that the **Jacobian operator** of a minimal Σ is defined as

$$L_\Sigma = \Delta_\Sigma + (\mathrm{Ric}(\nu,\nu) + |\mathrm{I\!I}|^2).$$

We say a minimal surface is **stable** if its second variation is always non-negative and its **index** is defined as the number of negative eigenvalues of its Jacobian operator counted with multiplicity.

It is clear that the minimal surface is defined in spirit of the geodesic. Each definition above is an analogue of a concept in the geodesic. The philosophy is also the same. That is, the geometry hide behind a FAMILY of objects. In classical Riemannian geometry, we have seen how to use Jacobian fields and second variation formulae to prove comparison theorems(see, for instance, [18]). Similarly, below we will see how a family of surfaces control the geometry of manifolds. For example, assuming the existence of an area-minimizing surface, we will prove some inequalities and determine the metric of certain manifolds.

Interestingly, there is some natural connection between the scalar curvature and minimal surfaces. This can be seen from the following Gauss equation: given $X, Y \in T_p\Sigma$, we have

$$\mathrm{Sec}^\Sigma|X\wedge Y|^2 - \mathrm{Sec}^M(X,Y)|X\wedge Y|^2 = \langle \mathrm{I\!I}(X,X), \mathrm{I\!I}(Y,Y)\rangle - \langle \mathrm{I\!I}(X,Y), \mathrm{I\!I}(X,Y)\rangle.$$

Summing over an orthonormal basis of $T_p\Sigma$, we get

$$\mathrm{Ric}(\nu,\nu) + |\mathrm{I\!I}|^2 = \frac{1}{2}(R^M - R^\Sigma + |H|^2 + |\mathrm{I\!I}|^2). \tag{1.2}$$

Note that the left two terms appears naturally in the second variation. This formula, which is known as Yau's rearrangement technique, is often seen in the study of minimal surfaces and also in the sequel. We will focus on 3 dimension below. In this case, $\frac{1}{2}R^\Sigma$ reduces to the Gauss curvature of Σ and Gauss-Bonnet formula gives a good interpretation on the topology. We will omit the superscript of R in this case.

Indeed, the philosophy behind the consideration is as follows. When we consider the sectional and Ricci curvature conditions, we use tools from geodesics. These two things are just like the weights on one side of a balance. To keep the equilibrium of the balance, once we weaken the curvatures to the scalar curvature, we have to strengthen geodesics to minimal surfaces.

There were numerous global results related to the scalar curvature in past decades. One may refer to Gromov's lecture notes [23] to get a full picture on the subject. In particular, many rigidity statements were found. That is, given some conditions involving the scalar curvature, we are able to determine the manifold. The most distinguished outcome among these statements should be the positive mass theorem. It claims that if a metric on a 3-manifold behaves like the flat metric on $\mathbb{R}^3$ at infinity and has non-negative scalar curvature, then the so-called ADM mass is non-negative. Moreover, this manifold will be isometric to $\mathbb{R}^3$ if the ADM mass vanishes. For the results on that direction, one may refer to a survey by Brendle [9].

1.3 Organization of The Paper

In this paper we will mainly focus on two aspects of the subject. First in section 2 we will collect several inequalities involving the scalar curvature and minimal surfaces. Here we will first use the Hersch's trick to establish a upper bound of the product of the area of a minimal surface and the minimum of the scalar curvature. This is a collection of results from Bray, Brendle, Neves [8], Ros [43] and Yau [48]. Using this estimate, combing some existence result, we can estimate the minimum area of surfaces of certain topology type given a lower bound of the scalar curvature [7, 8, 31]. We also collect a recent result given by Lowe and Neves on counting minimal surfaces in a negatively curved space using the tools from lamination and Ricci flow [30].

Then in section 3 and 4 we establish several rigidity results. In section two we concentrate on splitting results, which claims a product manifold given a least area surface. Here we discuss least area tori by Cai and Galloway [10, 11], least area spheres by Bray, Brendle, Neves [8] and least area hyperbolic surfaces by Nunes[37]. The main analysis tool used here is the constant mean curvature equation. At the end of the section we prove a warped product of hyperbolic manifold [3]. In section 3 we consider the rigidity case of the inequalities in section 1 and prove the rigidity results [7, 31, 30]. There the rigidity occurs when the metric has constant curvature. We will see in section 3 how to use the Ricci flow to find metric with constant curvature.

2 The Control of Geometric Quantities

In this section we will see how geometric quantities can be controlled by the scalar curvature. M will always denote a compact 3-manifold with a Riemannian metric g and Σ a compact surface. The area with respect to the metric g will be denoted by $\mathrm{Area}(\Sigma, g)$ or $\mathrm{Area}(\Sigma)$ for short if there is no confusion.

First, we use the Hersch's trick to establish several estimates for the area of minimal surfaces. This method was first used by Hersch in [25] to show that the round metric on $\mathbb{S}^2$ maximize the product of the first eigenvalue and the area. It is a clever way to choose test functions as the coordinates of a conformal map. The original proof by Hersch is a beautiful balancing argument which will also be seen in the following.

Proposition 2.1 ([8, 31, 43, 48], ref. [52]). *Let Σ be a compact embedded minimal surface with genus $g(\Sigma)$. Suppose $\inf_M R > 0$.*

(1) If Σ is stable and orientable, then Σ is a sphere with

$$\mathrm{Area}(\Sigma) \inf_M R \leqslant 8\pi.$$

(2) If Σ is stable and non-orientable, then

$$\operatorname{Area}(\Sigma)\inf_M R \leqslant 12\pi + 4\pi g(\tilde{\Sigma}),$$

where $\tilde{\Sigma}$ is the orientable double cover of Σ.

(3) If Σ is orientable and has index one, then

$$\operatorname{Area}(\Sigma)\inf_M R \leqslant 24\pi + 16\pi\left(\frac{g(\Sigma)}{2} - \left[\frac{g(\Sigma)}{2}\right]\right),$$

where $[x]$ denotes the integer part of x.

Note that the left sides of the inequalities have to be in the form of product due to the scaling invariance.

Proof. (1) Taking $\varphi = 1$ in (1.1) and by (1.2) we have

$$\int_\Sigma (R - 2K + |\mathrm{II}|^2)d\sigma = 2\int_\Sigma (\operatorname{Ric}(\nu,\nu) + |\mathrm{II}|^2)d\sigma \leqslant 0.$$

By Gauss-Bonnet formula,

$$0 < \operatorname{Area}(\Sigma)\inf_M R \leqslant \int_\Sigma R d\sigma \leqslant 2\int_\Sigma K d\sigma = 8\pi(1 - g(\Sigma)).$$

Hence $g(\Sigma) = 0$ and the desired results comes.

(2) Again we only need to show that $\int_\Sigma (\operatorname{Ric}(\nu,\nu) + |\mathrm{II}|^2)d\sigma \leqslant 4\pi(g(\tilde{\Sigma}) + 1))$. If Σ is two sided then, similar to (1), $\int_\Sigma \operatorname{Ric}(\nu,\nu) + |\mathrm{II}|^2 d\sigma \leqslant 0$. If Σ is one-sided, then the gerenal second variation formula reads

$$\int_\Sigma |(DX)^\perp|^2 - (\operatorname{Ric}(\nu,\nu) + |\mathrm{II}|^2)|X|^2 d\sigma \geqslant 0,\ \forall X \in N\Sigma, \tag{1.3}$$

where $N\Sigma$ is the normal bundle of Σ. Let $\tilde{\Sigma}$ be the two sided orientable double cover of Σ, $\pi : \tilde{\Sigma} \to \Sigma$ be the projection and $i : \tilde{\Sigma} \to \tilde{\Sigma}$ be the antipodal map. A vector field X on $\tilde{\Sigma}$ projects to a vector field on Σ if and only if $X \circ i = X$. Now Hersch's trick comes. By uniformization theorem, we can find a conformal map $\varphi : \Sigma \to \mathbb{RP}^2$. We may lift φ to a conformal map $\tilde{\varphi} : \tilde{\Sigma} \to \mathbb{S}^2 \subset \mathbb{R}^3$ and hence write $\tilde{\varphi} = (\tilde{\varphi}_1, \tilde{\varphi}_2, \tilde{\varphi}_3)$. Clearly $\tilde{\varphi}_\alpha \circ i = -\tilde{\varphi}_\alpha, \forall \alpha = 1,2,3$. Take a unit section $\tilde{V} \in \Gamma(\pi^* N\Sigma)$, then $\tilde{V} \circ i = -\tilde{V}$. Hence $\tilde{\varphi_\alpha}\tilde{V}$ projects to a normal vector filed V_α on Σ. Note that $\sum_{\alpha=1}^3 |V_i|^2 = 1$. Taking $X = V_\alpha$ in (1.3) and summing up,

$$\begin{aligned}\int_\Sigma (\operatorname{Ric}(\nu,\nu) + |\mathrm{II}|^2)d\sigma &\leqslant \sum_{\alpha=1}^3 \frac{1}{2}\int_{\tilde{\Sigma}} |\nabla^{\tilde{\Sigma}}\tilde{\varphi}_\alpha|^2 d\sigma_g \\ &= \frac{1}{2}\int_{\tilde{\Sigma}} |D\tilde{\varphi}|^2 d\sigma_{\tilde{\varphi}^* g_0} = \operatorname{Area}(\tilde{\varphi}(\tilde{\Sigma})) = \deg(\tilde{\varphi})\operatorname{Area}(\mathbb{S}^2),\end{aligned}$$

where we used the conformal invariance of the Dirichlet energy $\frac{1}{2}\int_{\tilde{\Sigma}} |D\tilde{\varphi}|^2$. Finally, by [33] we can choose φ such that $\deg\tilde{\varphi} \leqslant (g(\Sigma) + 1)$ which lead to the estimate.

(3) Similarly we only need to estimate $\int_\Sigma (\operatorname{Ric}(\nu,\nu) + |\mathrm{II}|^2)d\sigma$. Let $u > 0$ be an eigenfunction of the first eigenvalue of L_Σ. By spectral theory(see [21, Apendix D.6]),

$$\int_\Sigma |\nabla^\Sigma f|^2 - (\operatorname{Ric}(\nu,\nu) + |\mathrm{II}|^2)f^2 d\sigma \geqslant 0,\ \forall f \in C^\infty(\Sigma) \text{ with } \int_\Sigma f u d\sigma = 0.$$

Next we choose a conformal map $\hat{\varphi} : \Sigma \to \mathbb{S}^2$ that is orthogonal to u. In fact, first choose a conformal map $\varphi : \Sigma \to \mathbb{S}^2$ such that $\deg(\varphi) \leqslant \left[\frac{g(\Sigma)+1}{2}\right] + 1$ (Brill-Noether theory, see [22, P261]). Given $x \in \mathbb{S}^2$, let $\pi_x : \mathbb{S}^2 \setminus \{x\} \to \mathbb{C}$ be the stereographic projection. Since π_x is conformal, we have a family of conformal diffeomorphisms parameterized by $(x, t) \in \mathbb{S}^2 \times [0, 1)$,

$$\Psi(x, t) : \mathbb{S}^2 \to \mathbb{S}^2, y \mapsto \pi_x^{-1}\left(\frac{1}{1-t}\pi_x(y)\right).$$

Observe that $\Psi(x, 0)$ is independent of x, hence we can view Ψ as a continuous map from $\mathbb{B}^3$ to the group of conformal transformations of $\mathbb{S}^2$. Set

$$\Phi : \mathbb{B}^3 \to \mathbb{B}^3, v \mapsto \frac{1}{\int_\Sigma u d\sigma}\int_\Sigma (\Psi(v) \circ \varphi) u d\sigma.$$

Now, since $\lim_{y\to x, t\to 1^-} \Psi(y, t) = x, \forall x \in \mathbb{S}^2$, Φ can be continuously extended to $\bar{\mathbb{B}}^3$ with $\Phi|_{\mathbb{S}^2} = \mathrm{Id}_{\mathbb{S}^2}$. By Borsuk-Ulam Theorem, there exists a $v \in \mathbb{B}^3$ such that $\Phi(v) = 0$. $\hat{\varphi} = \Psi(v) \circ \varphi$ is the desired conformal map. Finally, arguing like (2) we get

$$\int_\Sigma (\mathrm{Ric}(\nu, \nu) + |\mathrm{II}|^2) d\sigma \leqslant 8\pi \deg(\hat{\varphi}) \leqslant 8\pi\left(\left[\frac{g(\Sigma)+1}{2}\right] + 1\right).$$

□

If $\min_M R < 0$, then the argument of (1) yields

Proposition 2.2 (I. Nunes, [37]). *If $\min_M R < 0$ and $\Sigma \subset M$ is orientable, minimal and stable, then*

$$-\mathrm{Area}(\Sigma) \inf_M R \geqslant 8\pi(g(\Sigma) - 1).$$

Combining with the existence results, we have two direct corollaries.

Definition 2.3. *(1) Let $\mathscr{F}$ be the collection of embedded surfaces $\Sigma \subset M$ such that Σ is homeomorphic to $\mathbb{RP}^2$. If $\mathscr{F} \neq \emptyset$, let*

$$\mathscr{A}(M, g) := \inf_{\Sigma\in\mathscr{F}} \mathrm{Area}(\Sigma, g).$$

(2) Similarly, let $\overline{\mathscr{F}}$ be the collection of smooth maps $\varphi : \mathbb{S}^2 \to M$ such that φ is non-trivial in $\pi_2(M)$. If $\pi_2(M) \neq 0$, let

$$\bar{\mathscr{A}}(M, g) := \inf_{\varphi\in\overline{\mathscr{F}}} \mathrm{Area}(\mathbb{S}^2, \varphi^* g).$$

Note that $\varphi^* g$ is just a pseudo-metric if φ is not an immersion, but we can still consider the area taken with respect to it as usual. Now,

Theorem 2.4 (H. Bray, S. Brendle, M. Eichmair, A. Neves [7]). *Suppose $\mathscr{F} \neq \emptyset$ and $R_g \geqslant 6$, then*

$$\mathscr{A}(M, g) \leqslant 2\pi.$$

Proof. Choose a minimizing sequence $\Sigma_k \in \mathscr{F}$ such that $\lim_{k\to\infty} \mathrm{Area}(\Sigma_k, g) = \mathscr{A}(M, g)$ and in particular $\limsup_{k\to\infty}(\mathrm{Area}(\Sigma_k, g) - \inf_{\Sigma\in\mathscr{I}(\Sigma_k)} \mathrm{Area}(\Sigma, g)) \leqslant 0$ where $\mathscr{I}(\Sigma)$ denotes the isotopic class of Σ. By [34, Theorem 1], there exists a subsequence of $\{\Sigma_k\}$ converges weakly to a disjoint union of stable embedded minimal surfaces. Since the construction in [34] is a genus non-increasing process, we know that one of the components of the resulting surface is homeomorphic to $\mathbb{RP}^2$. Consequently, there exists $\Sigma \in \mathscr{F}$ such that $\mathrm{Area}(\Sigma, g) = \mathscr{A}(M, g)$. By Theorem 2.1(2), we have the desired estimate. □

Remark. *Let $i : \Sigma \to M$ be the embedding. Then $i_* : \pi_1(\Sigma) \to \pi_1(M)$ is injective. Hence by [40, Theorem 1], we have* $\mathrm{Area}(\Sigma, g) \geqslant \mathrm{Sys}(\Sigma, g) \geqslant \mathrm{Sys}(M, g) > 0$*, where* $\mathrm{Sys}(\Sigma, g) := \inf\{L(\gamma) : \gamma \text{ is non-trivial in } \pi_1(\Sigma)\}$*. Hence $\mathscr{A}(M, g) > 0$, which ensures that we can choose the desired minimizing sequence.*

Theorem 2.5 (H. Bray, S. Brendle, A. Neves [8])**.** *Suppose $\pi_2(M) \neq 0$ and $R_g \geqslant 2$, then*

$$\bar{\mathscr{A}}(M, g) \leqslant 4\pi.$$

Proof. By [35, Theorem 7] there exists a stable minimal immersion $\varphi \in \overline{\mathscr{F}}$ which attains $\bar{\mathscr{A}}(M, g)$. Clearly the proof of Theorem 2.1 (1) still works for this situation and hence we have the estimate of $\bar{\mathscr{A}}(M, g)$. □

Recently, Theorem 2.5, and also the rigidity phenomenon we will state in the next section, has been generalized to the cases where M is an open manifold and where $3 \leqslant \dim M \leqslant 7$ by Zhu [49]. The main obstruction for the proof is that the existence and compactness results are no more available for the open manifold. The author therefore considered a modified area functional and then used the existence result for this functional to complete the proof.

Note that the standard model for Theorem 2.5 is not the $\mathbb{S}^2 \subset \mathbb{S}^3$ but $\mathbb{S}^2 \subset \mathbb{S}^2 \times \mathbb{R}/\Gamma$ for some appropriate group Γ. On the other hand, the standard $\mathbb{S}^2 \subset \mathbb{S}^3$, the equator, is not area-minimizing but a min-max surface.

The study of min-max surfaces also traces back to geodesics. Birkhoff first used sweepouts to find closed geodesics on $\mathbb{S}^2$ [5]. For the construction of geodesics, one may refer to [17, Chapter 5]. Later, Almgren [1] and Pitts [39] used the same idea to find minimal surfaces in closed manifolds. We briefly review the definition.

Let I be a closed interval and $\{\Sigma_t\}_{t\in I}$ be a family of closed subset of M with finite $\mathcal{H}^2$ measure. Roughly speaking, we say $\{\Sigma_t\}_{t\in I}$ is a **generalized family of surfaces**(or **sweepout**) if except at finite time and finite points of M they are surfaces and vary smoothly. An example is given by the level sets of the height function on $\mathbb{S}^3$. A family of sweepouts is **saturated** if it is invariant under smooth isotopies. For details, see [31].

Definition 2.6. *Given a saturated set Λ of M, the* ***width of M associated with*** Λ *is defined as*

$$W(M, \Lambda, g) := \inf_{\{\Sigma_t\}\in\Lambda} \sup_{t\in I} \mathcal{H}^2(\Sigma_t).$$

Recall also that an orientable embedded surface $\Sigma \subset M$ is a **Heegard splitting** if $M \setminus \Sigma$ is a disjoint union of two handle bodies. The **Heegard genus** of M is the lowest possible genus of such Σ. Given $h \in \mathbb{N} \cup \{0\}$, we denote by $\mathcal{E}_h$ the collection of all connected embedded minimal surfaces $\Sigma \subset M$ with $g(\Sigma) \leqslant h$. We say (M, g) satisfies the $(\star)_h$ condition if M contains no non-orientable embedded surface and each surface in $\mathcal{E}_h$ is unstable. Note that if $\mathrm{Ric}_g > 0$, then (M, g) admits no stable minimal surface.

If Σ is a Heegard splitting, then there is a natural collection of sweepouts associated with Σ. Each such sweepout $\{\Sigma_t\}_{t\in[-1,1]}$ satisfies

- $\Sigma_0 = \Sigma$ and Σ_t is an embedded surface isotopic to Σ for $t \in (-1, 1)$;
- $\{\Sigma_t\}_{-1\leqslant t\leqslant 0}$, $\{\Sigma_t\}_{0\leqslant t\leqslant 1}$ foliate two components of $M \setminus \Sigma$ respectively with Σ_{-1}, Σ_1 being graphs.

The saturated set generated by this collection is denoted by Λ_Σ. The union of all Λ_Σ such that Σ is a Heegard splitting of genus h is denoted by Λ^h. Our result is as follows:

Theorem 2.7 (Marques, Neves [31]). *Let h be the Heegard genus of M. Suppose* $\mathrm{Ric}_g > 0$ *and M admits no non-orientable embedded surface. If $R_g \geqslant 6$, then*

$$W(M, \Lambda^h, g) \leqslant 4\pi - 2\pi \left[\frac{h}{2}\right] \leqslant 4\pi.$$

By the work of Simon and Smith, for any saturated set Λ, there exists a min-max sequence converges in the varifold sense to an embedded smooth minimal surface which attains the width associated with Λ. For the construction, see [16]. However, this existence theorem is not enough for our use. We need some further information of the min-max surface.

Proposition 2.8. *Let h be the Heegard genus of M. Suppose (M, g) satisfies the $(\star)_h$ condition, then there exists an orientable embedded minimal surface Σ_0 with $g(\Sigma_0) = h$ such that*

$$\mathrm{Area}(\Sigma_0) = \inf_{\Sigma \in \mathcal{E}_h} \mathrm{Area}(\Sigma) = W(M, \Lambda^h).$$

Moreover, Σ_0 is index one and is contained in a sweepout $\{\Sigma_t\}_{t\in[0,1]} \in \Lambda_{\Sigma_0}$ with

(a) The function $f(t) = \mathrm{Area}(\Sigma_t)$ has a strict global maximum at $t = 0$,

(b) $f(t)$ is smooth around $t = 0$ with $f''(0) < 0$.

Proof. By [39], M contains at least one embedded minimal surface. Since h is also the lowest possible genus of embedded minimal surface in M, by standard compactness theorem(for instance, [4, Theorem 4.2]), there exists an embedded minimal Σ_0 with genus h such that $\mathrm{Area}(\Sigma_0) = \inf_{\Sigma\in\mathcal{E}_h} \mathrm{Area}(\Sigma)$. Next we construct a sweepout satisfying (a) (b).

By [34, Corollary of Theorem 5], each $\Sigma \in \mathcal{E}_h$ is a Heegard splitting. Denote by N_1, N_2 two components of $M \setminus \Sigma_0$. Let $\lambda < 0$ be the lowest eigenvalue of L_{Σ_0}, $u > 0$ be the corresponding eigenfunction and X a vector field on M with $X|_{\Sigma_0} = \varphi\nu$, where ν is a unit normal vector pointing into N_1. Denote by F_t the flow generated by X and $\Sigma_t = F_t(\Sigma_0)$. Then

$$\frac{\partial}{\partial t}\langle H(\Sigma_t), \nu\rangle\Big|_{t=0} = -L_{\Sigma_0} u = \lambda\varphi < 0.$$

Hence for $0 < t < \varepsilon$ small Σ_t is contained in N_1, the mean curvature vector points outward N_1 and $\mathrm{Area}(\Sigma_t) < \mathrm{Area}(\Sigma_0)$. Σ_ε bounds a handlebody N. We consider a sweepout $\{\tilde{\Sigma}_t\}_{t\in[\varepsilon,1]}$ of N such that $\tilde{\Sigma}_t = \Sigma_t$ for $t - \varepsilon$ small . Let $\tilde{\Lambda}$ be the saturated set of N generalized by $\{\tilde{\Sigma}_t\}$. Note that $\partial N = \Sigma_\varepsilon$. We must have $W(N, \tilde{\Lambda}) \leqslant \mathrm{Area}(\partial N)$. Otherwise since $H(\partial N) > 0$, $W(N, \tilde{\Lambda})$ can be achieved by a minimal surface in the interior of N with genus h, and thus disjoint from Σ_0. However, every two surfaces in $\mathcal{E}_h$ must intersect with each other, a contradiction. Hence by $\mathrm{Area}(\partial N) < \mathrm{Area}(\Sigma_0)$ and definition, we can find a $\{\Sigma_t\}_{t\in[0,\varepsilon]} \in \tilde{\Lambda}$ so that $\sup_{\varepsilon\leqslant t\leqslant 1} \mathrm{Area}(\Sigma_t) < \mathrm{Area}(\Sigma)$. Similarly we can define $\{\Sigma_t\}_{t\in[-1,0]}$.

Now we have a sweepout $\{\Sigma_t\}_{t\in[-1,1]}$ satisfying (a). For (b), we have

$$f''(0) = -\int_{\Sigma_0} u L_{\Sigma_0} u \, d\sigma = \lambda \int_{\Sigma_0} u^2 d\sigma < 0.$$

By (a) (b), Σ_0 is index one. Finally, by definition $W(M, \Lambda^h) \leqslant \mathrm{Area}(\Sigma_0)$. Suppose on the contrary that $W(M, \Lambda^h) < \mathrm{Area}(\Sigma_0)$, then by [29, Theorem 0.6], we can find an embedded minimal surface Σ' such that $\mathrm{Area}(\Sigma') = W(M, \Lambda^h)$ and $g(\Sigma') \leqslant g(\Sigma_0)$. This contradicts the fact that $\mathrm{Area}(\Sigma_0) = \inf_{\Sigma\in\mathcal{E}_h} \mathrm{Area}(\Sigma)$. □

We remark that later Xin Zhou showed that the min-max hypersurface in a manifold with positive Ricci curvature is either orientable and index one or stable and non-orientable [50] [51]. Also, Marques and Neves showed that the min-max surface with k parameters has index no more than k [32].

Proposition 2.1 (3) and Proposition 2.8 is still not enough for Theorem 2.7. We use a blow up technique to improve the estimate. To do this, consider the Ricci flow on M:

$$\begin{cases} \dfrac{\partial g}{\partial t}(t) = -2\mathrm{Ric}_{g(t)}, & t \in [0, T); \\ g(0) = g. \end{cases} \tag{1.4}$$

Lemma 2.9. *Let h be the Heegard genus of M. If $(M, g(t))$ satisfies the $(\star)_h$ condition for $0 \leqslant t < T'$, then*

$$W(M, \Lambda^h, g(t)) \geqslant W(M, \Lambda^h, g) - \left(16\pi - 8\pi \left[\frac{h}{2}\right]\right) t, \quad \forall\, 0 \leqslant t < T'.$$

Proof. Let $T'' < T$ and suppose $\sup_M |\mathrm{Ric}_{g(t)}|_{g(t)} \leqslant C$ for all $t \in [0, T'']$, then $e^{-2C|t_1 - t_2|} g(t_1) \leqslant g(t_2)$ for all $t_1, t_2 \in [0, T'']$. Thus $e^{-2C|t_1 - t_2|} W(M, \Lambda^h, g(t_1)) \leqslant W(M, \Lambda^h, g(t_2))$, which makes $W(M, \Lambda^h, g(t))$ locally Lipschitz. In particular $W(M, \Lambda^h, g(t))$ is differentiable almost everyhere on $t \in [0, T)$.

Let $t_0 \in [0, T')$ be a time such that $W(M, \Lambda^h, g(t))$ is differentiable at t_0. By Proposition 2.8, there exists a sweepout $\{\Sigma_s\}_{s \in [-1.1]}$ satisfying (a) (b) in Proposition 2.8 with Σ_0 index one. We consider $f(s,t) := \mathrm{Area}(\Sigma_s, g(t))$ which is smooth around $(0, t_0)$ and satisfies $\frac{\partial f}{\partial s}(0, t_0) = 0$, $\frac{\partial^2 f}{\partial s^2}(0, t_0) < 0$. Consequently by implicit function theorem there exists a smooth function $s(t)$ with $s(t_0) = 0$ such that $(s(t), t)$ is the unique solution for $\frac{\partial f}{\partial s}(s, t) = 0$ near $(0, t_0)$. On the other hand, since $s \mapsto f(s, t_0)$ has a strict global maximum at $s = 0$, $s = s(t)$ is also the global strict maximum of $s \mapsto f(s, t)$ for $|t - t_0|$ small. In particular,

$$\sup_{s \in [-1,1]} \mathrm{Area}(\Sigma_s, g(t)) = \mathrm{Area}(\Sigma_{s(t)}, g(t)) \text{ for } |t - t_0| \text{ small,}$$

which makes $\sup_{s \in [-1,1]} \mathrm{Area}(\Sigma_s, g(t))$ is smooth around $t = t_0$. Now,

$$W(M, \Lambda^h, g(t)) \leqslant \sup_{s \in [-1,1]} \mathrm{Area}(\Sigma_s, g(t)) \text{ for all } t \text{ and } W(M, \Lambda^h, g(t_0)) = \sup_{s \in [-1,1]} \mathrm{Area}(\Sigma_s, g(t_0))$$

$$\sup_{s \in [-1,1]} \mathrm{Area}(\Sigma_s, g(t)) \geqslant \mathrm{Area}(\Sigma_0, g(t)) \text{ for all } t \text{ and } \sup_{s \in [-1,1]} \mathrm{Area}(\Sigma_s, g(t_0)) = \mathrm{Area}(\Sigma_0, g(t_0))$$

Hence we have

$$\frac{d}{dt} W(M, \Lambda^h, g(t)) \bigg|_{t = t_0} = \frac{d}{dt} \sup_{s \in [-1,1]} \mathrm{Area}(\Sigma_s, g(t)) \bigg|_{t = t_0} = \frac{d}{dt} \mathrm{Area}(\Sigma_0, g(t)) \bigg|_{t = t_0}.$$

Using (1.2), Proposition 2.1 (3) and Ricci flow equation (1.4), we get

$$\begin{aligned} \frac{d}{dt} \mathrm{Area}(\Sigma_0, g(t))\big|_{t = t_0} &= -\int_{\Sigma_0} ((R_{g(t_0)} - \mathrm{Ric}_{g(t_0)}(\nu, \nu))) d\sigma_{g(t_0)} \\ &= -4\pi \mathcal{X}(\Sigma_0) - \int_{\Sigma_0} (\mathrm{Ric}_{g(t_0)}(\nu, \nu) + |\mathrm{II}_g(t_0)|^2_{g(t_0)}) d\sigma_{g(t_0)} \\ &\geqslant 8\pi(h - 1) - 8\pi \left(\left[\frac{h+1}{2} \right] + 1 \right) = -16\pi + 8\pi \left[\frac{h}{2} \right]. \end{aligned}$$

As a result, $\frac{d}{dt} W(M, \Lambda^h, g(t)) \geqslant -16\pi + 8\pi \left[\frac{h}{2}\right]$ a.e. $t \in [0, T')$ and the desired estimate comes. □

Proof of Theorem 2.7. Let $g(t), t \in [0,T)$ be a maximal solution of the Ricci flow (1.4) on M. By [24], $\mathrm{Ric} > 0$ is preserved by the Ricci flow and hence $\mathrm{Ric}_{g(t)} > 0$ for all t. Thus $(M, g(t))$ satisfies the $(\star)_h$ condition for all t, which makes

$$W(M, \Lambda^h, g(t)) \geqslant W(M, \Lambda^h, g) - \left(16\pi - 8\pi\left[\frac{h}{2}\right]\right)t, \quad \forall 0 \leqslant t < T.$$

By Proposition 2.1 (3), $W(M, \Lambda^h, g(t)) \min_M R_{g(t)} \leqslant 24\pi + 16\pi\left(\frac{h}{2} - \left[\frac{h}{2}\right]\right)$ for all $t \in [0,T)$. Also, by [24, Theorem 15.1], $\lim_{t\to T} \min_M R_{g(t)} = +\infty$. Hence $\lim_{t\to T} W(M, \Lambda^h, g(t)) = 0$. Taking limit, we get

$$W(M, \Lambda^h, g) \leqslant \left(16\pi - 8\pi\left[\frac{h}{2}\right]\right)T.$$

On the other hand, we have the evolution equation of the scalar curvature

$$\frac{\partial}{\partial t} R_{g(t)} = \Delta_{g(t)} R_{g(t)} + 2|\mathrm{Ric}_{g(t)}|^2_{g(t)} \geqslant \Delta_{g(t)} R_{g(t)} + \frac{2}{3} R^2_{g(t)}.$$

By maximum principle, we have $R_g(t) \geqslant \frac{1}{1-4t}$. Hence $T \leqslant \frac{1}{4}$. As a result, we get the desired estimate. $\square$

We end this section by remarking a most recent result in this direction. In the following we assume M is a hyperbolic manifold, i. e. admits a metric g_0 with constant sectional curvature -1. Recall that there exists a discrete and torsion-free subgroup $\Gamma < \mathrm{PSL}(2,\mathbb{C}) = \mathrm{Isom}^+(\mathbb{H}^3)$ such that $M = \mathbb{H}^3/\Gamma$. Denote by the $\mathbb{S}^2_\infty$ the **sphere at infinity**(think of the topological boundary of the Poincare model $\mathbb{B}^3 \subset \mathbb{R}^3$). Let $\overline{\mathbb{H}}^3 := \mathbb{H}^3 \cup \mathbb{S}^2_\infty$, endowed with the cone topology(see [2]). Given a set $\Omega \subset \mathbb{H}^3$, we denote by $\overline{\Omega}$ the closure of Ω in $\overline{\mathbb{H}}^3$ and $\partial_\infty \Omega := \overline{\Omega} \cap \mathbb{S}^2_\infty$. The **limit set** $\Lambda(G) \subset \mathbb{S}^2_\infty$ for a discrete subgroup $G < \mathrm{PSL}(2,\mathbb{C})$ acting properly discontinuously on $\mathbb{H}^3$ is defined as the set of accumulation points in $\mathbb{S}^2_\infty$ of the orbit Gx, which is independent of $x \in \mathbb{H}^3$.

A C^1 map $F : \mathbb{S}^2_\infty \to \mathbb{S}^2_\infty$ is said to be $(1+\varepsilon)$**-quasiconformal** for some $\varepsilon > 0$ if for all p, DF_p sends circles into ellipses whose eccentricity is bounded by $1+\varepsilon$. If $\Lambda(\Gamma)$ is a Jordan curve and there exists an $(1+\varepsilon)$-quasiconformal map which maps the equator to $\Lambda(\Gamma)$, then $\Lambda(\Gamma)$ is called an $(1+\varepsilon)$**-quasicircle**. The set of all $(1+\varepsilon)$-quasicircle is denoted by C_ε.

Given a closed immersed surface $i : \Sigma \to M$. If $i_* : \pi_1(\Sigma) \to \pi_1(M) = \Gamma$ is injective, then it is said to be **essential**. Hence we can associate its limit set $\Lambda(\Sigma) := \Lambda(\pi_1(\Sigma)) \subset \mathbb{S}^2_\infty$.

Definition 2.10. *Denote by $S_\varepsilon(M)$ the set of all homotopy class Π of essential surface with limit set in C_ε.*
(1) Define

$$\mathrm{Area}(\Pi, g) := \inf_{\Sigma\in\Pi} \mathrm{Area}(\Sigma, g).$$

*(2) The **minimal surface entropy** is defined as*

$$E(g) := \lim_{\varepsilon\to 0} \limsup_{L\to\infty} \frac{\ln \#\{\Pi \in S_\varepsilon(M) : \mathrm{Area}(\Pi, g) \leqslant 4\pi(L-1)\}}{L \ln L}.$$

The minimal surface entropy was first studied by Calegari, Marques and Neves. They showed that if $\mathrm{Sec}_g \leqslant -1$, then $E(g) \geqslant 2 = E(g_0)$ in [12]. On the contrary, we have the upper bound of $E(g)$:

Theorem 2.11 (B. Lowe, A. Neves [30])**.** *If $R_g \geqslant -6$, then*

$$E(g) \leqslant 2.$$

3 Rigidity Phenomena I: Splitting Results

Next we characterize the cases where the equalities hold. In this section we present several splitting results. That is, we determine the situation where a given manifold (M^3, g) can be written as the product $M = \Sigma^2 \times \mathbb{R}$ endowed with the product metric $g = g_\Sigma + dt$, where Σ is an embedded surface in M. A classical argument to prove the result in this form is finding a global parallel vector field, which can be seen in the Cheeger-Gromoll splitting theorem ([13, 19], see also [18]) and De Rham decomposition theorem ([42], see also [6, p.288]). This idea is natural because $\partial/\partial t$ is such a vector field for the product $\Sigma \times \mathbb{R}$. In the following, this vector field is realized by the normal vector of the surfaces in a foliation.

In the previous section we minimized the area of spheres and projective planes in 3-manifold. One may ask whether there is a similar result to Theorem 2.4 and Theorem 2.5 for tori. But the answer is probable no. The standard embedding for a torus is $\mathbb{T}^2 \to \mathbb{T}^3$, where the ambient space is flat. Hence we cannot expect any upper bound for the product $\mathrm{Area}(\mathbb{T}^2) \min_M R$. On the other hand, only assuming a non-negative scalar curvature and the existence of least area torus, we have the following result.

Definition 3.1. *Given an embedded surface Σ in M and a neighborhood U of Σ, we denote by $\mathscr{I}(\Sigma, U)$. $\mathscr{I}(\Sigma)$ the isotopy class of Σ in U and in M respectively.*

(1) We say Σ is ***area-minimizing*** *if Σ is of least area in $\mathscr{I}(\Sigma)$,*

(2) we say Σ is ***locally area-minimizing*** *if Σ is of least area in $\mathscr{I}(\Sigma, U)$ for some neighborhood U of Σ.*

Theorem 3.2 (Cai, Galloway [10, 11])**.** *Suppose $R_g \geqslant 0$ and the boundary of M (possibly empty) is mean convex. If M contains a two-sided torus Σ which is area-minimizing, then M is covered by $I \times \mathbb{T}^2$, where $I = [0, l]$ or $\mathbb{R}$ and the metric on $\mathbb{T}^2$ is flat.*

The authors first proved Theorem 3.2 when M is analytic in [10] and later when M is smooth in [11]. In order to prove Theorem 3.2, we only need to prove the following local result: from local area-minimizing to local splitting. Then Theorem 3.2 follows from a routine continuous argument.

Theorem 3.3. *Let M be as in Theorem 3.2. If an embedded two-sided torus Σ of M is locally area-minimizing, then a neighborhood of Σ is isometric to $\Sigma \times (-\varepsilon, \varepsilon)$.*

Proof. Using normal exponential map, there is a neighborhood U of Σ such that the metric on U can be written as

$$g = dt^2 + \sum_{i,j=1}^{2} g_{ij}(x,t) dx_i \otimes dx_j, \quad x \in \Sigma,\ t \in (-\varepsilon, \varepsilon).$$

When M is analytic, the theorem is a direct consequence by inductively computing

$$\frac{\partial^n g_{ij}}{\partial t^n}(x, 0) = 0, \quad \forall n = 1, 2, \dots,\ i, j = 1, 2 \text{ and } x \in \Sigma.$$

See [10]. Note that the case $n = 1$ is from

$$\int_\Sigma |\mathrm{II}|^2 d\sigma \leqslant \int_\Sigma (|\mathrm{II}|^2 + R) d\sigma \leqslant \int_\Sigma K d\sigma = 0.$$

When M is just smooth, a few techniques are needed. First we modify g to $\hat{g}$ whose component $\hat{g}_{ij}(t, x) := g_{ij}(t^2/|t|, x)$ is even. We next show that Σ cannot be strict area-minimizing in $(U, \hat{g})$.

Choose ε small such that Σ is of least area in $\mathscr{I}(\Sigma, U)$. If Σ were strict of least area in $\mathscr{I}(\Sigma, U)$, we consider $\hat{g}_n := e^{-2n^{-1}t^2}\hat{g}$ is a metric on U which satisfies $R_{\hat{g}_n} > 0$ when n large and ε small enough, and converges to g in the C^∞ topology. Choose a smooth positive function $f(t)$ on $(-\varepsilon, \varepsilon)$ such that $f|_{[-\frac{\varepsilon}{2}, \frac{\varepsilon}{2}]} = 0$ and $|f'(\pm\varepsilon)|$ large enough with $f'(\varepsilon) > 0, f'(-\varepsilon) < 0$, then the boundary of U is mean convex with respect to $\bar{g} := e^f\hat{g}, \bar{g}_n := e^f\hat{g}_n$. By [34, Theorem 1'], there exists $\Sigma_n \in \mathscr{I}(\Sigma)$ such that

$$\operatorname{Area}(\Sigma, \bar{g}_n) = \alpha_n := \inf_{\Sigma' \in \mathscr{I}(\Sigma, U)} \operatorname{Area}(\Sigma', \bar{g}_n).$$

For each n, Σ is a compact stable minimal surface with respect to $\bar{g}_n$ and $\{\alpha_n\}$ is bounded. As a result, after passing to a subsequence, Σ_n converges to a compact minimal torus $\bar{\Sigma}$ with respect to $\bar{g}$. Then $\bar{\Sigma}$ is isotopic to Σ since each Σ_n is. Furthermore, we have

$$\operatorname{Area}(\bar{\Sigma}, \bar{g}) \leqslant \liminf_{n\to\infty} \operatorname{Area}(\Sigma_n, \bar{g}_n) \leqslant \liminf_{n\to\infty} \operatorname{Area}(\Sigma, \bar{g}_n) \leqslant \operatorname{Area}(\Sigma, \bar{g}).$$

But Σ is also strict of least area in $(U, \bar{g})$. Hence $\bar{\Sigma} = \Sigma$. By the convergence, Σ_n contains in $t \in (-\frac{\varepsilon}{2}, \frac{\varepsilon}{2})$ where $\bar{g}_n = \hat{g}_n$ which yields Σ_n is of least area in $(U, \hat{g}_n)$ for n large. But this is impossible due to the second variation formula and $R_{\hat{g}_n} > 0$.

Hence there exists $\bar{\Sigma}$ which is also of least area in $(U, \bar{g})$ and distinct from Σ. Then $\bar{\Sigma}$ cannot intersect with Σ otherwise they will meet transversally because they are both totally geodesic. By reflecting the portion of $\bar{\Sigma}$ in $t < 0$ to $t > 0$ and smoothing, we get a surface which is isotopic to Σ but has less area, a contradiction.

The above argument shows that there are two minimal torus Σ^+, Σ^-, one on each side of Σ, each is isotopic to Σ and locally of least area. Let $V \subset U$ be the region bounded by Σ^+ and Σ^-. Then V is topological $[-1, 1] \times \mathbb{T}^2$ by the property of isotopy. Gluing two W along their boundary, we get a manifold with non-negative scalar curvature which is diffeomorphic to $\mathbb{T}^3$ and hence, by [46], is flat. Finally, since Σ is totally geodesic, it is also flat. Hence V is isometric to $[-1, 1] \times \mathbb{T}^2$. □

Proof of Theorem 3.2. Without loss of generality we may assume Σ is a boundary component of M, other cases can be easily reduces to this case. By the previous lemma, for some maximal $l > 0$, $\Phi : \Sigma \times [0, l) \to M, (x, t) \mapsto \exp_x t\nu$ is an isometry. If $l = \infty$, then $\Phi(\Sigma \times [0, \infty))$ is both open and closed in M and hence equals to M. Next we assume $l < \infty$. Clearly Φ extends to $\Sigma \times [0, l]$. If $\Phi : \Sigma \times [0, l] \to M$ is isometric, then $\Phi(\Sigma \times \{l\})$ is a boundary component of M and hence $\Phi(\Sigma \times [0, l]) = M$.

Were $\Phi : \Sigma \times [0, l] \to \Phi(\Sigma \times [0, l])$ not an isometry. Then for some distinct $x_1, x_2 \in \Sigma$, we have $\Phi(x_1, l) = \Phi(x_2, l)$ and there is some neighborhood U_i of x_i so that $\Phi|_{U_i \times [0,l]}$ is an isometry. Since $\Phi(U_i \times \{l\}), i = 1, 2$ are two totally geodesic surfaces, they must agree near $\Phi(x_i, l)$. Hence $\Phi(\Sigma \times [0, l])$ is both closed and open, which must be M. □

Recently there is a similar result to Theorem 3.2 for the embedded cylinder [15]. The proof requires a careful insight of the topology of the solution to certain Plateau problems.

Next we consider the rigidity cases of Proposition 2.2 and Theorem 2.5. Again we want to construct a foliation near Σ where each surface is area-minimizing. There is a uniform way to construct this foliation via a family of constant mean curvature developed in [3, 8]. The argument we show below can also be adapted to give an alternative proof of Theorem 3.2, see [36].

Proposition 3.4 ([8, 37]). *Let Σ be a compact orientable two-sided immersed surface in M. If $\mathrm{II}, \mathrm{Ric}(\nu,\nu)$ vanishes along Σ, then there is an $\varepsilon > 0$ and a smooth function w on $\Sigma \times (-\varepsilon, \varepsilon)$ such that*

(a) For all $x \in \Sigma$, we have $w(x,0) = 0$ and $\frac{\partial}{\partial t} w(x,t)\big|_{t=0} = 1$,

(b) for all $t \in (-\varepsilon,\varepsilon)$, we have $\int_\Sigma (w(\cdot,t) - t) d\sigma = 0$,

(c) for all $t \in (-\varepsilon,\varepsilon)$, the surfaces

$$\Sigma_t = \{\exp_x(w(x,t)\nu(x)) : x \in \Sigma\}$$

has constant mean curvature.

Proof. By the assumption, the Jacobian operator of Σ is $L_\Sigma = \Delta_\Sigma$. For fixed $\alpha \in (0,1)$, we consider two Banach spaces $X := \{u \in C^{2,\alpha}(\Sigma) : \int_\Sigma u d\sigma = 0\}$ and $Y := \{u \in C^{0,\alpha}(\Sigma) : \int_\Sigma u d\sigma = 0\}$ and $\Sigma_u := \{\exp_x(u(x)\nu(x)) : x \in \Sigma\}$ for $u \in X$. Choose $\delta > 0$ such that Σ_{u+t} is a compact $C^{2,\alpha}$ surface for $t \in (-\delta,\delta), u \in B_\delta(0)$, where $B_\delta(0)$ is the unit ball centered at the origin of X. Finally, consider

$$\Psi : B_\delta(0) \times (-\varepsilon,\varepsilon) \to Y, (u,t) \mapsto H_{\Sigma_{u+t}} - \frac{1}{\mathrm{Area}(\Sigma)} \int_\Sigma H_{\Sigma_{u+t}} d\sigma.$$

Then $\Psi(0,0) = 0$ and the linearization of Ψ at $(0,0)$ is $D\Psi(0,0) = -L_\Sigma = -\Delta_\Sigma$, which is an isomorphism from X to Y. Hence by implicit function theorem, there exists $\varepsilon > 0$ and $u(t) = u(\cdot,t) \in B_\delta(0)$ for $t \in (-\varepsilon,\varepsilon)$ with

$$u(0) = 0 \text{ and } \Psi(u(t),t) = 0, \text{ for all } t \in (-\varepsilon,\varepsilon).$$

Hence $w(x,t) := u(x,t) + t$ satisfies the desired conditions. □

Theorem 3.5 (Rigidity statement of Theorem 2.5). *If $R_g \geqslant 2$ and $\bar{\mathscr{A}}(M,g) = 4\pi$, then M is covered by $\mathbb{S}^2 \times \mathbb{R}$.*

Proof. Again, we only need to prove the local splitting. By [35], there exists an immersion $\varphi : \mathbb{S}^2 \to M$ such that $\mathrm{Area}(\mathbb{S}^2, \varphi^* g) = 4\pi$. By the proof of Proposition 2.1, we have

$$\int_{\mathbb{S}^2} (\mathrm{Ric}(\nu,\nu) + |\mathrm{II}|^2) d\sigma_{\varphi^* g} = 0.$$

By the stability, 1 lies in the kernel of L_Σ which makes $\mathrm{Ric}(\nu,\nu) + |\mathrm{II}|^2 = 0$. Moreover,

$$\int_{\mathbb{S}^2} (R + |\mathrm{II}|^2) d\sigma_{\varphi^* g} = 0.$$

Together with $R_g \geqslant 2$ we have $R = 2$ and $|\mathrm{II}|^2 = 0$ on $\mathbb{S}^2$. By (1.2), the Gauss curvature $K_{\varphi^* g} = 1$ and $\varphi^* g = g_0$ is the round metric on $\mathbb{S}^2$. Next consider $w(x,t)$ given by Proposition 3.4 and let $\varphi_t(x) := \exp_{\varphi(x)}(w(x,t)\nu(x))$, $\Sigma_t := \varphi_t(\mathbb{S}^2)$. Let $\nu_t(x), \mathrm{II}_t, H(t)$ be the normal vector, second fundamental form and mean curvature of Σ_t respectively with $\nu_0 = \nu$. We want to show that $\mathrm{Area}(\mathbb{S}^2, f_t^* g) = 4\pi$ for small t.

To this end, consider the **lapse function** defined by $\rho_t(x) := \left\langle \nu_t(x), \frac{\partial}{\partial t}\varphi_t(x) \right\rangle$. By [26, equation (1.2)], $\Delta_{\varphi_t^* g}\rho_t + (\mathrm{Ric}(\nu_t,\nu_t) + |\mathrm{II}_t|^2)\rho_t = -H'(t)$. Set $\bar{\rho}_t$ be the average of ρ_t with respect to $\varphi^* g$. By Poincare's inequality and since $\mathrm{Ric}(\nu_t,\nu_t) + |\mathrm{II}_t|^2 \to 0$ uniformly as $t \to 0$, we have

$$\int_{\mathbb{S}^2} |\nabla^{\varphi_t^* g}\rho_t|^2_{\varphi_t^* g} - (\mathrm{Ric}(\nu_t,\nu_t) + |\mathrm{II}_t|^2)(\rho_t - \bar{\rho}_t)^2 d\sigma_{\varphi_t^* g} \geqslant 0 \text{ for } t \text{ small}.$$

On the other hand, since $\varphi^* g$ is of least area 4π in its isotopy class, we have

$$8\pi \leqslant \operatorname{Area}(\mathbb{S}^2, \varphi_t^* g) \inf_M R \leqslant 8\pi + \int_{\mathbb{S}^2} (\operatorname{Ric}(\nu_t, \nu_t) + |\mathrm{II}_t|^2) d\sigma_{\varphi_t^* g}.$$

Inserting into the last two inequality and with the equation of ρ_t, we have

$$\begin{aligned} 0 &\leqslant \int_{\mathbb{S}^2} |\nabla^{\varphi_t^* g} \rho_t|^2 + (\operatorname{Ric}(\nu_t, \nu_t) + |\mathrm{II}_t|^2) \rho_t (2\bar{\rho}_t - \rho_t) d\sigma_{\varphi_t^* g} \\ &= -H'(t) \int_{\mathbb{S}^2} (2\bar{\rho}_t - \rho_t) d\sigma_{\varphi_t^* g} = -H'(t) \int_{\mathbb{S}^2} \rho_t \sigma_{\varphi_t^* g}. \end{aligned}$$

Since $\rho_0 = 1$, we have $\rho_t > 0$ for t small and hence $H'(t) \leqslant 0$. This leads that $H(t) \leqslant 0$, which makes $\frac{d}{dt} \operatorname{Area}(\mathbb{S}^2, \varphi_t^* g) = H(t) \int_{\mathbb{S}^2} \rho_t d\sigma_{\varphi_t^* g} \leqslant 0$ if $t \geqslant 0$ small. Arguing similarly for $t < 0$, we have $\operatorname{Area}(\mathbb{S}^2, \varphi_t^* g) \leqslant \operatorname{Area}(\mathbb{S}^2, \varphi^* g) = 4\pi$. As a result, $\operatorname{Area}(\mathbb{S}^2, \varphi_t^* g) = 4\pi$ for t small. Hence $\varphi_t^* g$ also minimizes area and we have $\operatorname{Ric}(\nu_t, \nu_t) = 0, \mathrm{II}_t = 0$ for small t. The equation of ρ_t yields $\Delta_{\varphi_t^* g} \rho_t = 0$ on $\mathbb{S}^2$ which means that ρ_t is a constant. An easy calculation shows that ν_t is parallel for small t, and hence $\Phi : \mathbb{S}^2 \times \mathbb{R} \to M, (x, t) \to \exp_{\varphi(x)} t\nu(x)$ is a local isometry for t small.

Finally, by a routine continuation argument we have $\Phi : \mathbb{S}^2 \times \mathbb{R} \to M$ is a local isometry, and hence is a Riemannian covering map. □

Theorem 3.6 (Rigidity statement of Proposition 2.2)**.** *Suppose $R_g \geqslant -2$. If there exists an embedded orientable surface Σ which is area-minimizing and* $\operatorname{Area}(\Sigma) = 4\pi(g(\Sigma) - 1)$*, then M is covered by $(\Sigma \times \mathbb{R}, g_\Sigma + dt^2)$, where g_Σ is a hyperbolic metric.*

Proof. We point out the difference between the proof of Theorem 3.6 and that of Theorem 3.5. Again we have $\operatorname{Ric}(\nu, \nu), \mathrm{II} = 0, R = -2, K = -1$ along Σ. Let $w(x, t)$ be the function given by Proposition 3.4. Similarly define $H(t)$ as in the proof of Theorem 3.5. Let $\varphi : \Sigma \times (-\varepsilon, \varepsilon) \to M$ be $\varphi(x, t) := \exp_x w(x, t)\nu(x)$.

We prove that $H(t) \leqslant 0$ for $t \geqslant 0$ small. Suppose, on the contrary, that there exists ε_k decreases to 0 such that $H(\varepsilon_k) > 0$ for all k. Consider $V_k := \Sigma \times [0, \varepsilon_k]$ and a metric $g_k := (\varphi|_{V_k})^* g$ on V_k. Then $R_{g_k} \geqslant -2$ and (V_k, g_k) has mean convex boundary. Moreover, $\operatorname{Area}(\Sigma, g_k) = 4\pi(g(\Sigma) - 1)$. In order to get a contradiction, we recall some facts about the Yambabe equation(ref. [20]). The **Sobolev quotient** of a Riemannian manifold (N^n, h) is defined as

$$Q_h(N) := \inf_{f \in C^1(M) : f \neq 0} Q_h(f),$$

where

$$Q_h(f) := \frac{\int_N (|\nabla^h f|_h^2 + \frac{n-2}{4(n-1)} R_h f^2) d\mathrm{Vol}_h + \frac{n-2}{2(n-1)} \int_{\partial N} H_h f^2 d\sigma_h}{(\int_N |f|^{2n/(n-2)} d\mathrm{Vol}_h)^{(n-2)/n}}.$$

If $Q_h(N) < Q(\mathbb{S}^n_+)$, then there exists a smooth minimizer of $Q_h(N)$ and h is conformal to a metric with constant scalar curvature. Moreover, we can choose the metric such that the boundary of N is minimal.

We next show that $Q_{g_k}(V_k) < 0 \leqslant Q(\mathbb{S}^3_+)$ for k large. Indeed, since $R_g = -2$ on Σ, we have $-2 \leqslant R_{g_k} \leqslant -1$ for k large. Hence

$$Q_{g_k}(1) = \frac{\frac{1}{8} \int_{V_k} R_{g_k} d\mathrm{Vol}_{g_k} + \frac{1}{4} \int_{\partial V_k} H_{g_k} d\sigma_{g_k}}{\mathrm{Vol}(V_k, g_k)^{1/3}} \leqslant \frac{-\frac{1}{8} \mathrm{Vol}(V_k, g_k) + \frac{1}{4} H(\varepsilon_k) \operatorname{Area}(\Sigma_{\varepsilon_k}, g_k)}{\mathrm{Vol}(V_k, g_k)^{1/3}}.$$

Calculation gives the expansion $H(\varepsilon_k) = O(\varepsilon_k^2)$ and $\mathrm{Vol}(V_k, g_k) = \varepsilon_k \mathrm{Area}(\Sigma, g_k) + O(\varepsilon_k^2)$. As a result, $Q_{g_k}(1) < 0$ for k large enough. Thus, there exists $u \in C^\infty(V_k)$ such that $u^4 g_k$ has constant scalar curvature -2 and ∂V_k is minimal. This is equivalent to the fact that u solves

$$\begin{cases} \Delta_{g_k} u - \frac{1}{8} R_{g_k} u - \frac{1}{4} u^5 = 0 & \text{in } V_k \\ \frac{\partial u}{\partial \eta} + \frac{1}{4} H_{g_k} u = 0 & \text{on } \partial V_k. \end{cases}$$

Since $R_{g_k} \geqslant -2$, we have $\Delta_{g_k} u \geqslant \frac{1}{4}(u^5 - u)$ on V_k. By the maximum principle, we have $u < 1$ or $u = 1$. The latter is impossible because this means that $H(\varepsilon_k) = 0$, a contradiction. Hence $u < 1$, which make $\mathrm{Area}(\Sigma, u^4 g_k) < \mathrm{Area}(\Sigma, g_k) = 4\pi(g(\Sigma) - 1)$. By [34, Theorem 1'], we can minimize the area of surfaces in $\mathcal{I}(\Sigma, V_k)$ to get a stable minimal surface $\bar{\Sigma}$ with the genus of Σ. But this contradicts to Proposition 2.2 because $R_{u^4 g_k} = -2$ and $\mathrm{Area}(\bar{\Sigma}, u^4 g_k) < 4\pi(g(\bar{\Sigma}) - 1)$. Consequently, $H(t) \leqslant 0$ for $t \geqslant$ small.

Similarly $H(t) \geqslant 0$ for $t \leqslant 0$ small. Then we can repeat the argument in the proof of Theorem 3.5. □

For the hyperbolic space, note that $\mathbb{H}^3$ is a warped product of $\mathbb{R}^2 \rtimes \mathbb{R}, g = e^{2t}(dx^2 + dy^2) + dt^2$. To get the splitting result in this form, we need to consider a different functional. Let M be an $(n+1)$-dimensional orientable Riemannian manifold with volume form $\mathbf{\Omega}$. We assume there is a global defined n-form Λ such that $\mathbf{\Omega} = d\Lambda$. The **brane action** is defined as

$$\mathcal{B}(\Sigma) := \mathrm{Area}(\Sigma) - n \int_\Sigma \Lambda.$$

By Stokes formula, if Σ bounds some region U, then $\int_\Sigma \Lambda = \mathrm{Vol}(U)$. In some texts, the brane action is also referred to the n-bubbles, see [23, Chapter 5]. The first variation formula for the brane action related to the variation vector field $f\nu$ is

$$\mathcal{B}'(0) = \int_\Sigma (H - n) f \, d\sigma.$$

and the second variation is same to that of usual volume functional. Hence the critical points of $\mathcal{B}$ is surfaces with constant mean curvature n. We have an analogue of Theorem 3.2 for the brane action.

Theorem 3.7 (Andersson, Cai, Galloway [3])**.** *Let (M^{n+1}, g) be an orientable Riemannian manifold with $R_g \geqslant -n(n+1)$. If there exists a compact orientable hypersurface Σ in M which does not admit a metric of positive scalar curvature and locally minimizes the brane action, then a neighborhood U of Σ is isometric to the warped product $(\Sigma \times (-\varepsilon, \varepsilon), e^{2t} h + dt^2)$, where the induced metric h on Σ is Ricci flat.*

Proof. Let h be the induced metric on Σ. We first show that $R_g = -n(n+1), \mathrm{I\!I} = h$ and $\mathrm{Ric}_h = 0$ along Σ. Choose a smooth function u on Σ such that $u > 0$ and $L_\Sigma u \leqslant 0$. Then the scalar curvature of the conformal metric $\tilde{h} := u^{\frac{2}{n-2}} h$ is

$$R_{\tilde{h}} = u^{-\frac{2}{n-2}} \left(-2u^{-1} L_\Sigma u + R_g + n(n-1) + |\mathring{\mathrm{I\!I}}| + \frac{n-1}{n-2} \frac{|\nabla^\Sigma u|^2}{u^2} \right) \geqslant 0,$$

where $\mathring{\mathrm{I\!I}} = \mathrm{I\!I} - h$ is the trace-free second fundamental form. Were $R_{\tilde{h}} > 0$ at some points, then by [27] there were a metric with positive scalar curvature in the conformal class of $\tilde{h}$. This contradicts our assumption. Hence $R_{\tilde{h}}$ vanishes identically, which makes $\mathring{\mathrm{I\!I}} = 0, L_\Sigma u = 0, R_g = -n(n+1)$ and u is constant on Σ. By (1.2), $S^\Sigma = 0$. Hence Σ admits a metric with positive scalar curvature unless h is Ricci flat.

Hence by (1.2), L_Σ reduces to Δ_Σ. Similar to Proposition 3.4, we get a smooth $w(x,t)$ on Σ such that $\Sigma_t := \{\Phi(x,t) := \exp_x(w(x,t)\nu(x)) : x \in \Sigma\}$ has constant mean curvature $H(t)$ for t small. Hence under the coordinate given by Φ, there is a neighborhood $U = \Phi(\Sigma \times (-\varepsilon, \varepsilon))$ with $g|_U = h_t + \varphi^2 dt^2$. Similar to previous two theorems, we show that Σ is also the local maximum of the brane action. That is $H(t) \leqslant n$ for $t \geqslant 0$ small. Were the statement false, then there were a $t_0 > 0$ small with $H(t_0) > n$ and $H'(t_0) > 0$. The scalar curvature of the metric $\bar{h} := \varphi^{\frac{2}{n-2}} h_{t_0}$ on Σ_{t_0} is

$$R_{\bar{h}} = \varphi^{-\frac{2}{n-2}} \left(-2\varphi^{-1} H'(t_0) + R_g + |\mathrm{II}_{t_0}|^2 + H(t_0)^2 + \frac{n-1}{n-2} \frac{|\nabla^{\Sigma_{t_0}} \varphi|^2}{\varphi^2} \right) > 0,$$

because $|\mathrm{II}_{t_0}|^2 \geqslant H^2/n > n$. This is a contradiction because Σ_{t_0} is diffeomorphic to Σ for t small. Similarly we have $H(t) \geqslant n$ for $t \leqslant 0$ small. As a result, for these small t, Σ_t also minimize the brane action locally and hence $H(t) = n$. Moreover, $\Delta_{\Sigma_t} \varphi = L_{\Sigma_t} \varphi = -H'(t) = 0$, which makes φ is a constant on each Σ_t. By variational principle, each Σ_t is stable under the brane action and hence $\mathrm{II}_t = h_t$ for t small. Finally, by a change of t-coordinate in $g|_U = h_t + \varphi^2 dt^2$, we may assume $\varphi = 1$. Then $\mathrm{II}_t = h_t$ becomes

$$\frac{\partial h_t}{\partial t} = 2h_t.$$

Integration gives that $h_t = e^{2t} h$, which is the desired form of the metric on U. □

The global version of Theorem 3.7 is the following special case of the positive mass theorem for the hyperbolic space.

Theorem 3.8 (Andersson, Cai, Galloway [3]). *Suppose* $(M^n, g), 3 \leqslant n \leqslant 7$ *has scalar curvature* $R_g \geqslant -n(n-1)$ *and is isometric to* $\mathbb{H}^n$ *outside a compact set. Then* (M, g) *is globally isometric to* $\mathbb{H}^n$.

The proof uses a construction of a minimizing sequence of the brane action, which requires a little caution.

4 Rigidity phenomena II: finding constant curvature via the Ricci flow

As we have seen in the first section, the Ricci flow is defined as the solution to

$$\begin{cases} \dfrac{\partial g}{\partial t}(t) = \mathrm{Ric}_{g(t)}; \\ g(0) = g. \end{cases}$$

The Ricci flow was first studied by Hamilton in [24], where he proved the short time existence and uniqueness of the Ricci flow using the DeTurck trick. Also he showed that the positive Ricci curvature is preserved by the Ricci flow in 3-dimension. With the time going, the Ricci flow turns out to be a useful tool in solving certain geometry problems. As said Gromov in his lecture notes [23], the **solutions to elliptic equations** are

kinds of residues of certain geometric or topological complexity of the underlying manifolds, that is necessary for the very existence of these solutions,

while the **Ricci flow**,

as a road roller, leaves a flat terrain behind itself as is it crawls along erasing all kinds of complexities.

In Theorem 2.7 we have seen how the Ricci flow reduced difficulties in our problems. The motivation of Hamilton to develop the Ricci was to prove the celebrated Poincare's conjecture, which is equivalent to find a metric with constant curvature on a closed simply connected 3-manifold. Hamilton proved this assertion on a positive Ricci manifold and later the general case was solved by Perelman. In the following we will also see how the Ricci flow will help us to find a metric with constant curvature on a given manifold.

Theorem 4.1 (Rigidity statement of Theorem 2.4). *If $R_g \geqslant 6$ and $\mathscr{A}(M,g) = 2\pi$, then (M,g) is isometric to $\mathbb{RP}^3$.*

Proof. Let $g(t), 0 \leqslant t < T$ be a maximal solution of the Ricci flow on M with $g(0) = g$. Similar to Theorem 2.7, we have

$$\mathscr{A}(M, g(t)) \geqslant \mathscr{A}(M,g) - 8\pi t = 2\pi(1-4t).$$

Since Theorem 2.4 yields $\mathscr{A}(M,g(t)) \inf_M R_{g(t)} \leqslant 12\pi$, we have $\inf_M R_{g(t)} \leqslant \frac{1}{1-4t}$. On the other hand, by the evolution equation of the scalar curvature

$$\frac{\partial}{\partial t} R_{g(t)} = \Delta_{g(t)} R_{g(t)} + 2|\mathrm{Ric}_{g(t)}|^2_{g(t)} \geqslant \Delta_{g(t)} R_{g(t)} + \frac{2}{3} R^2_{g(t)}$$

together with the maximum principle, we have

$$\inf_M R_{g(t)} = \frac{1}{1-4t}.$$

By the strong maximum principle, $R_{g(t)} = \frac{1}{1-4t}$ on M for $t \in [0,T)$. By the evolution of the scalar curvature again, we know that the trace-free part of $\mathrm{Ric}_{g(t)}$ vanishes (note $\mathrm{Ric} = \mathring{\mathrm{Ric}} + \frac{R}{3} g$) and, in particular, 3-manifold $(M, g(t))$ has constant curvature for all t.

Finally we only need to show that $\#\pi_1(M) = 2$. Let $\Sigma \in \mathscr{F}$ which attains $\mathscr{A}(M,g)$. Using the local isometry from $\mathbb{S}^3 \to M$, we can easily show that $i_* : \pi_1(\Sigma) \to \pi_1(M)$ is surjective. Together with the fact that i_* is injective, we know $\pi_1(M) = \mathbb{Z}_2$ and hence M is isometric to $\mathbb{RP}^3$. ◻

Theorem 4.2 (Rigidity statement of Theorem 2.7). *Let M, h be as in Theorem 2.7. If $R_g \geqslant 6$ and $W(M, g, \Lambda^h) = 4\pi$, then M is isometric to $\mathbb{S}^3$.*

Proof. Since $W(M, \Lambda^h, g) = 4\pi$, by the proof of Theorem 2.7, we must have $h = 0, 1$. Similar to the proof of Theorem 4. 1, we have M is of constant curvature 1. We next show that $h = 0$.

If $h = 1$, there is an embedded minimal torus T which realizes the width. By Proposition 2.8, the area of other minimal tori is bigger than 4π. However, the classical classification theorem [44, Theorem 1.7] yields that the manifolds with Heegard genus one are either the lens space $L(p,q)$ or $\mathbb{S}^2 \times \mathbb{S}^1$, which contains a flat torus of area $2\pi^2/p < 4\pi$, which is a contradiction. Hence by [44, Theorem 1.4], M is isometric to $\mathbb{S}^3$. ◻

Similarly, we can characterize the case where the equality holds in Theorem 2.11.

Theorem 4.3. *If $R_g \geqslant -6$ and $E(g) = 2$, then g has constant curvature -1.*

Up to now, we have seen several results which provide a lot of insights for the scalar curvature. However, they are not satisfying. On the one hand, most quantities we considered do not have a clear geometric interpretation. On the other hand, our results highly rely on the theorems that is only available in 3-dimension, such as the uniformization theorem and Gauss-Bonnet formula. When the dimension turns higher, the cases will become much more complicated and is still undiscovered. This is, just as Gromov said, The mystery of the scalar curvature remains unsolved!

Reference

[1] F. Almgren, *The Theory of Varifolds*, Mimeographed notes, 1965.
[2] M. Anderson, Complete minimal varieties in hyperbolic space, *Inventiones Mathematicae*, 69, 477-494, 1982.
[3] L. Andersson, M. Cai and G. Galloway, Rigidity and positivity of mass for asymptotically hyperbolic manifolds *Annales Henri Poincaré*, 9(1), 1-33, 2008.
[4] M. Anderson, Curvature estimates for minimal surfaces in 3-manifolds, *Annales scientifiques de l'École Normale Supérieure*, 18(1), 89-105, 1985.
[5] G. D. Birkhoff, Dynamical systems with two degrees of freedom, *Transaction of American Mathematical Society*, 18(2), 199-300, 1917.
[6] A. Besse, *Einstein Manifolds*, Springer Science & Business Media, 2007.
[7] H. Bray, S. Brendle, M. Eichmair and A. Neves, Area-minimizing projective planes in 3-manifolds *Communications on Pure and Applied Mathematics*, 63(9), 1237-1247, 2010.
[8] H. Bray, S. Brendle and A. Neves, Rigidity of area-minimizing two-spheres in three-manifolds, *Communications in Analysis and Geometry*, 18(4), 821-830, 2010.
[9] S. Brendle, Rigidity phenomena involving scalar curvature, *Surveys in Differential Geometry, volume XVII*, 179-202, 2012.
[10] M. Cai and G. Galloway, Least area tori and 3-manifolds of non-negative scalar curvature, *Mathematische Zeltschrlft*, 223, 387-395, 1996.
[11] M. Cai and G. Galloway, Rigidity of area minimizing tori in 3-manifolds of non-negative scalar curvature, *Communications in Analysis and Geometry*, 8(3), 565-573, 2000.
[12] D. Calegari, F. Marques, A. Neves. Counting minimal surfaces in negatively curved 3-manifolds, *Duke Mathematical Journal*, 171(8), 1615-1648, 2022.
[13] J. Cheeger and D. Gromoll, The splitting theorem for manifolds of nonnegative Ricci curvature, *Journal of Differential geometry*, 6(1), 119-128, 1971.
[14] B. Chen and L. Vanhecke, Total curvature of geodesic spheres, *Archiv der Mathematik*, 32(1), 404-411, 1979.
[15] O. Chodosh, M. Eichmair and V. Moraru, A splitting theorem for scalar curvature, *Communications on Pure and Applied Mathematics*, 72(2), 1231-1242, 2019.
[16] T. Colding and C. De Lellis, The min-max construction of minimal surfaces, *Surveys in Differential Geometry VIII*, International Press, 75-107, 2003.
[17] T. Colding and W. Minicozzi II, *A Course in Minimal Surfaces*, Graduated Studies in Mathematics, vol. 121, American Mathematical Society, 2011.
[18] X. Dai and G. Wei, *Comparison Geometry for Ricci Curvature*, web. math. ucsb. edu/ dai/Ricci-book. pdf, 2012.
[19] J. Eschenburg and E. Heintze, An elementary proof of the Cheeger-Gromoll splitting theorem, *Annals of Global Analysis and Geometry*, 2(2), 1984.
[20] J. Escobar, The Yamabe problem on manifolds with boundary, *Journal of Differential Geometry*, 35(1), 21-84, 1992.
[21] L. Evans, *Partial Differential Equations, Second Edition*, Graduate Studies in Mathematics, vol. 19, American Mathematical Society, 2010.
[22] P. Griffiths and J. Harris, *Principles of Algebraic Geometry*, Wiley-Interscience series, 1978.
[23] M. Gromov, *Four Lectures on Scalar Curvature*, arXiv:1908. 10612 [math. DG], 2021.
[24] R. Hamilton, Three-manifolds with positive Ricci curvature, *Journal of Differential geometry*, 17(2), 1982.
[25] J. Hersch, Quatre propriétés isopérimétriques de membranes sphériques homogènes, *Comptes rendus de l'Académie des Sciences, Série A*, 270(5), 1645-1648, 1970.
[26] G. Huisken and T. Ilmanen, The inverse mean curvature flow and the Riemannian Penrose inequality, *Journal of Differential Geometry*, 59(3), 353-437, 2001.
[27] J. Kazdan and F. Warner, Existence and conformal deformation of metrics with prescribed Gaussian and scalar curvatures, *Annals of Mathematics*, 101(2), 317-331, 1975.
[28] D. Knopf and A. Young, Asymptotic stability of the cross curvature flow at a hyperbolic metric, *Proceeding of the American Mathematical Society*, 137(2), 699-709, 2009.
[29] C. De Lellis and F. Pellandini, Genus bounds for minimal surfaces arising from min-max constructions, *Journal für die reine und angewandte Mathematik*, 644, 47-99.
[30] B. Lowe and A. Neves, Minimal Surface Entropy and Average Area Ratio, arXiv:2110. 09451 [math. DG], 2021.
[31] F. Marques and A. Neves, Rigidity of min-max minimal spheres in three-manifolds, *Duke Mathematical Journal*, 161(14), 2725-2752, 2012.
[32] F. Marques and A. Neves, Morse index and multiplicity of min-max minimal hypersurfaces, *Cambridge Journal of Mathematics*, 4(4), 463-511, 2016.
[33] G. Martens, Minimale Blatterzahl bei Überlagerungen Kleinscher Flachen über der projektiven Ebene, *Archiv der Mathematik*, 30, 481-486, 1978.

[34] W. Meeks, L. Simon and S. T. Yau, Embedded minimal surfaces, exotic spheres, and manifolds with positive Ricci curvature, *Annals of Mathematics*, 116(3), 621-659, 1982.

[35] W. Meeks and S. T. Yau, Topology of three dimensional manifolds and the embedding problems in minimal surface theory, *Annals of Mathematics*, 112(3), 441-481, 1980.

[36] M. Micallef and V. Moraru, Splitting of 3-manifolds and rigidity of area-minimizing surfaces, *Proceedings of the American Mathematical Society*, 143(7), 2865-2872, 2015.

[37] I. Nunes, Rigidity of Area-Minimizing hyperbolic surfaces in three-manifolds, *Journal of Geometric Analysis*, 23(3), 1290-1302, 2013.

[38] G. Perelman, Ricci flow with surgery on three-manifolds, arXiv:0303109 [math. DG], 2003.

[39] J. Pitts, *Existence and Regularity of Minimal Surfaces on Riemannian Manifolds*, Mathematical Notes 27, Princeton University Press, 1981.

[40] P. M. Pu, Some inequalities in certain nonorientable Riemannian manifolds, *Pacific Journal of Mathematics*, 2(1), 55-71, 1952.

[41] M. Ratner, Raghunathan’s topological conjecture and distributions of unipotent flows, *Duke Mathematical Journal*, 63(1), 235-280, 1991.

[42] G. De Rham, Sur la réductibilité d’un espace de Riemann, *Commentarii Mathematici Helvetici*, 26, 328-344, 1952.

[43] M. Ross, The second variation of nonorientable minimal submanifolds, *Transaction of The American Mathematical Society*, 349(8), 3093-3104, 1997.

[44] N. Saveliev, *Lectures on the Topology of 3-Manifolds : an Introduction to the Casson Invariant, 2nd Edition*, de Gruyter, 2011.

[45] R. Schoen and S. T. Yau, On the proof of the positive mass conjecture in general relativity, *Communications in Mathematical Physics*, 65(1), 45-76, 1979.

[46] R. Schoen and S. T. Yau, On the structure of manifolds with positive scalar curvature, *Manuscripta Mathematica*, 28, 159-183, 1979.

[47] A. Seppi, Minimal discs in hyperbolic space bounded by a quasicircle at infinity, *Commentarii Mathematici Helvetici*, 91(4), 807-839, 2016.

[48] S. T. Yau, Nonlinear analysis in geometry, *L’ Enseignement Mathématique*, 33, 109-158, 1987.

[49] J. Zhu, Rigidity results for complete manifolds with nonnegative scalar curvature, arXiv:2008. 07028 [math. DG] 2020.

[50] X. Zhou, Min-max minimal hypersurface in (M^{n+1}, g) with $\mathrm{Ric}_g > 0$ and $2 \leqslant n \leqslant 6$, *Journal of Differential Geometry*, 100(1), 129-160, 2015.

[51] X. Zhou, Min-max hypersurface in manifold of positive Ricci curvature, *Journal of Differential Geometry*, 105(2), 291-343, 2017.

[52] X. Zhou, *Lecture Notes on Minimal Surfaces*, drive.google.com/file/d/109aB5rqRfwEJlOnVQGZyG2H3 66dz0HA0/view, 2021.


笔墨诗篇

稻花香里说丰年，听取蛙声一片。

——《西江月 · 夜行黄沙道中》

词四首

张文锦

踏莎行

露润流光，风摇轻影，桂华满枕凉撄梦。
那堪趸泣渍竹痕，灯花坠尽清宵永。
萤灭疏烟，星销幽暝。霜枫敲落绝魂声。
鲛珠抛碎忍捐玦？晓寒暗锁愁千重。

破阵子

茄管为谁凄哽？琼花似我沉浮。
曾许昆山寻片玉，独曳青藜探戴丘。峥嵘莫比俦。
可供几重鲲曝，平添多少怊惆。
谩道繁嚣沤沫事，且挂箕瓢守旧幽。何劳身外求？

一剪梅

雨掷秋声作浅凉，露沐霞枝，云醮霜江。
雁衔远恨欲谁招？梦绕闲潭，冷彻西窗。
撷萃落萤试饮香，揉桂为诗，雅韵深藏。
飘泠风月谢纷华，恬淡浮名，拚却清狂。

西江月・樱

风驭巧编绡素，羲和善浣香胭。
彤云偏爱降人间，撩惹诗情一片。
纤雨招邀玉蕊，共趋林下翩跹。
多情落粉尚缠绵，掷弃芳心谁怨？

作者信息：张文锦，2021 级数学科学学院，邮箱：ygnzbddfbx@mail.ustc.edu.cn.
稿件信息：2023 年 2 月 9 日来稿，3 月 7 日接受，稿件编辑：吴佳骏，卢昊翔。

皇榜

稻花香里说丰年，听取蛙声一片。

——《西江月 · 夜行黄沙道中》

皇榜第三期第 2 题解答

考虑 n 阶随机矩阵 $M_n = (e_{ij})_{n\times n}$, 其中 $\mathbb{P}(e_{ij} = 1) = \mathbb{P}(e_{ij} = -1) = 1/2$, 且所有分量独立. 本题的目的是证明一个关于估计 $\det M_n$ 的猜想.

(1) 证明: $\mathbb{E}[(\det M_n)^2] = n!$;

(2) 证明: 对于任意的函数 $f(n)$, 若 $\lim_{n\to\infty} f(n) = \infty$, 则 $\lim_{n\to\infty} \mathbb{P}\left(|\det M_n| \le f(n)\sqrt{n!}\right) = 1$;

(3) 设 X 是 M_n 的第 1 列, W 是 M_n 的第 $2, \ldots, d+1$ 列张成的线性子空间, 其中 $1 \le d \le n-4$, n 充分大. 记 $d(X, W)$ 表示 n 维 Euclid 空间中向量 X 对应的点到子空间的 W 的距离. 证明:

$$\mathbb{P}\left(|d(X,W) - \sqrt{n-d}| \ge t+1\right) \le 4\exp(-t^2/16), \quad \forall t > 0;$$

(4) 证明: $\lim_{n\to\infty} \mathbb{P}\left(|\det M_n| \ge g(n)\sqrt{n!}\right) = 1$, 其中 $g(n) = \exp\left(-29\sqrt{n\log n}\right)$;

(5) 证明: $|\det M_n| = n^{n(1/2-o(1))}$ a.s. 成立, 其中 $o(1)$ 表示 $n \to \infty$ 时的无穷小量.

本题由刘党政老师供题.

解答:

(1) 设 S_n 是 $\{1, 2, \cdots, n\}$ 的置换群, 利用

$$\det(M_n) = \sum_{\sigma\in S_n} \mathrm{sgn}(\sigma) \prod_{i=1}^n e_{i\sigma(i)},$$

有

$$\begin{aligned}
\mathbb{E}[(\det(M_n))^2] &= \sum_{\sigma\in S_n}\sum_{\sigma'\in S_n} \mathrm{sgn}(\sigma)\mathrm{sgn}(\sigma')\mathbb{E}\left[\prod_{i=1}^n e_{i\sigma(i)}e_{i\sigma'(i)}\right] \\
&= \sum_{\sigma\in S_n}\sum_{\sigma'\in S_n} \mathrm{sgn}(\sigma)\mathrm{sgn}(\sigma')\prod_{i=1}^n \mathbb{E}[e_{i\sigma(i)}e_{i\sigma'(i)}] \\
&= \sum_{\sigma\in S_n} (\mathrm{sgn}(\sigma))^2 \prod_{i=1}^n \mathbb{E}[e_{i\sigma(i)}^2] = \#\{\sigma\in S_n\} \\
&= n!.
\end{aligned}$$

(2) 利用 Markov 不等式, 结合 (1) 得

$$\mathbb{P}\left(|\det(M_n)| > f(n)\sqrt{n!}\right) = \mathbb{P}\left((\det(M_n))^2 > f^2(n)n!\right) \le \frac{\mathbb{E}[(\det(M_n))^2]}{f^2(n)n!} = \frac{1}{f^2(n)} \xrightarrow{n\to+\infty} 0,$$

即

$$\lim_{n\to+\infty} \mathbb{P}\left(|\det(M_n)| \le f(n)\sqrt{n!}\right) = 1.$$

(3) 只需证明对任意固定的子空间 W, 待证命题均成立. 设矩阵

$$A := \begin{pmatrix} e_{21} & e_{31} & \cdots & e_{d+1,1} \\ e_{22} & e_{32} & \cdots & e_{d+1,2} \\ \vdots & \vdots & \cdots & \vdots \\ e_{2n} & e_{3n} & \cdots & e_{d+1,n} \end{pmatrix},$$

记 $P=A(A^TA)^{-1}A^T, P=(p_{ij})_{n\times n}$, 利用最小二乘法可知, PX 为向量 X 在子空间 W 上的投影, 且投影矩阵 P 满足

$$P^T=P,\quad P^2=P,\quad \mathrm{tr}(P)=d.$$

因此

$$d(X,W)^2=\|X\|^2-\|PX\|^2=n-X^TPX=n-d-X^TBX,$$

其中 $B=P-D$, D 为 P 的对角元组成的矩阵 $\mathrm{diag}(p_{11},\cdots p_{nn})$, B 为零对角元的实对称矩阵. 记 $B=(b_{ij})_{n\times n}$, 利用 $b_{ii}=0$ 知

$$\mathbb{E}[X^TBX]=\sum_{k=1}^n b_{kk}\mathbb{E}[e_{1k}^2]+\sum_{1\le i\ne j\le n} b_{ij}\mathbb{E}[e_{1i}e_{1j}]=0.$$

因此

$$\mathbb{E}[d(X,W)^2]=n-d.$$

现在我们先给出一个引理:

引理 1(Hoeffding 不等式). 设随机变量 $X_1,X_2,\cdots,X_n$ 独立, 且满足 $a_i\le X_i\le b_i$. 记 $S_n=X_1+X_2+\cdots+X_n$, 则对 $\forall x>0$, 有

$$\mathbb{P}(S_n-\mathbb{E}S_n\ge x)\le \exp\left\{-\frac{2x^2}{\sum_{i=1}^n(b_i-a_i)^2}\right\}.$$

类似地,

$$\mathbb{P}(|S_n-\mathbb{E}S_n|\ge x)\le 2\exp\left\{-\frac{2x^2}{\sum_{i=1}^n(b_i-a_i)^2}\right\}.$$

引理 1 证明: 只证第一个等式: 设 $\mathbb{E}X_i=0$, 对任意 $t>0$, 有

$$\mathbb{P}(S_n\ge x)\le \mathrm{e}^{-tx}\mathbb{E}\mathrm{e}^{tS_n}=\mathrm{e}^{-tx}\prod_{k=1}^n\mathbb{E}\mathrm{e}^{tX_k}. \tag{1.1}$$

下面我们估计 $\mathbb{E}\mathrm{e}^{tX_i}$. 取 $\gamma=\frac{x-a_i}{b_i-a_i}$, 利用 Jenson 不等式, 我们有

$$\mathrm{e}^{tx}:=f(x)=f(\gamma a_i+(1-\gamma)b_i)\le \gamma f(a_i)+(1-\gamma)f(b_i)\le \frac{x-a_i}{b_i-a_i}\mathrm{e}^{tb_i}+\frac{b_i-x}{b_i-a_i}\mathrm{e}^{ta_i}.$$

因此

$$\mathbb{E}\mathrm{e}^{tX_i}\le -\frac{a_i}{b_i-a_i}\mathrm{e}^{tb_i}+\frac{b_i}{b_i-a_i}\mathrm{e}^{ta_i}=(1-\theta+\theta\mathrm{e}^{t(b_i-a_i)})\mathrm{e}^{-\theta t(b_i-a_i)}=(1-\theta+\theta\mathrm{e}^u)\mathrm{e}^{-\theta u}=\mathrm{e}^{g(u)}.$$

其中

$$\theta=-\frac{a_i}{b_i-a_i},\ u=t(b_i-a_i),\ g(u)=-\theta u+\log(1-\theta+\theta\mathrm{e}^u).$$

而 $g(0)=g'(0)=0,\ g''(u)\le\frac14(\forall u>0)$. 利用 Taylor 展开, 存在 $\xi\in(0,u)$, 满足

$$g(u)=g(0)+g'(0)u+\frac{g''(\xi)}{2}u^2\le\frac{u^2}{8}=\frac{t^2(b_i-a_i)^2}{8}.$$

因此 $\mathbb{E}\mathrm{e}^{tX_i}\le \mathrm{e}^{\frac{t^2(b_i-a_i)^2}{8}}$. 将上式代入 (1.1) 得

$$\mathbb{P}(S_n\ge x)\le\exp\left\{-tx+\frac{t^2}{8}\sum_{i=1}^n(b_i-a_i)^2\right\}$$

取 $t=\frac{4x}{\sum_{i=1}^n(b_i-a_i)^2}$ 即得该不等式.

回到原题: 设 $Y_{ij} = b_{ij}e_{1i}e_{1j}$, 记 $S_n := \sum_{i,j=1}^{n} Y_{ij} = X^T BX$. 利用上述引理 (Hoeffding 不等式), 结合 $\mathbb{E}[S_n] = 0$, $-|b_{ij}| \le Y_{ij} \le |b_{ij}|$, 有

$$\begin{aligned}\mathbb{P}\left(|d(X,W)-\sqrt{n-d}| \ge t+1\right) &= \mathbb{P}\left(|d^2(X,W)-(n-d)| \ge (t+1)|d(X,W)+\sqrt{n-d}|\right)\\ &\le \mathbb{P}\left(|S_n| \ge (t+1)\sqrt{n-d}\right)\\ &\le 2\exp\left\{-\frac{(t+1)^2(n-d)}{2\sum_{i,j=1}^{n} b_{ij}^2}\right\}.\end{aligned} \tag{1.2}$$

而

$$d = \sum_{k=1}^{n} p_{kk} = \mathrm{tr}(P) = \mathrm{tr}(P^2) = \sum_{i,j=1}^{n} p_{ij}^2,$$

由 Cauchy 不等式知,

$$\sum_{i,j=1}^{n} b_{ij}^2 = \sum_{i,j=1}^{n} p_{ij}^2 - \sum_{k=1}^{n} p_{kk}^2 \le d - \frac{d^2}{n} \le n-d.$$

代入 (1.2) 得

$$\mathbb{P}(\ |d(X,W)-\sqrt{n-d}| \ge t+1) \le 2\exp\left\{-\frac{(t+1)^2(n-d)}{2\sum_{i,j=1}^{n} b_{ij}^2}\right\} \le 2\exp\left\{-\frac{(t+1)^2}{2}\right\} \le 4\exp\left\{-\frac{t^2}{16}\right\}.$$

注: 原题干条件 “$1 \le d \le n-4$” 可省去, 只需 “$1 \le d \le n-1$” 即可.

(4) 设 W_i 是 M_n 的第 $1,2,\cdots,i$ 列生成的线性子空间, X_i 是 M_n 的第 i 列, 利用行列式的几何意义可知

$$|\det(M_n)| = \|X_1\| \prod_{i=1}^{n-1} d(X_{i+1},W_i) = \sqrt{n}\prod_{i=1}^{n-1} d(X_{i+1},W_i).$$

对 $k_i > 0$, 注意到

$$\mathbb{P}(|\det(M_n)| < \sqrt{n}k_1k_2\cdots k_{n-1}) \le \sum_{i=1}^{n-1} \mathbb{P}(d(X_{i+1},W_i) < k_i),$$

目标 1: 找到合适的 k_i(可能与 n 有关), 使得 $\sqrt{n}k_1k_2\cdots k_{n-1} \ge g(n)\sqrt{n!}$, 且

$$\lim_{n\to+\infty} \sum_{i=1}^{n-1} \mathbb{P}(d(X_{i+1},W_i) < k_i) = 0. \tag{1.3}$$

现在, 我们要找到合适的 $l(n)$, 让两段分别处理且均收敛至 0. 一方面, 当 $i \le l(n)$ 时, 我们取 $k_i = \sqrt{n-i}-t_i-1$, $t_i = \sqrt{32\log(n-i)}$, 并利用 (3) 中的结论, 有

$$\sum_{i=1}^{l(n)} \mathbb{P}\left(d(X_{i+1},W_i) < k_i\right) \le \sum_{i=1}^{l(n)} \mathbb{P}\left(|d(X,W)-\sqrt{n-d}| > t_i+1\right) \le 4\sum_{i=1}^{l(n)} \exp\left\{-\frac{t_i^2}{16}\right\} = \sum_{i=1}^{l(n)} \frac{4}{(n-i)^2}. \tag{1.4}$$

由 (1.4) 知, 我们找的 $l(n)$ 只需满足 $n-l(n) \to +\infty$ 即可 (记为目标 2, 这里先假设 $l(n)$ 已找到). 同时,

$$\begin{aligned}\sqrt{n}k_1k_2\cdots k_{n-1} &= \frac{\sqrt{n!}}{\sqrt{(n-l(n)-1)!}} \prod_{i=1}^{l(n)}\left(1-\frac{t_i+1}{\sqrt{n-i}}\right)\prod_{j=l(n)+1}^{n-1} k_j\\ &\ge \frac{\sqrt{n!}}{\sqrt{(n-l(n)-1)!}} \exp\left(-2\sum_{i=1}^{l(n)}\frac{t_i+1}{\sqrt{n-i}}\right)\prod_{j=l(n)+1}^{n-1} k_j.\end{aligned} \tag{1.5}$$

其中 $0 < \dfrac{t_i+1}{\sqrt{n-i}} < \dfrac{1}{2}$ 保证上式最后一个不等号成立. 而

$$\sum_{i=1}^{l(n)} \frac{t_i+1}{\sqrt{n-i}} \leq 6\sum_{i=1}^{n-1}\sqrt{\frac{\log i}{i}} \leq 6\sqrt{\log n}\sum_{i=1}^{n-1}\frac{1}{\sqrt{i}} \leq 12\sqrt{n\log n},$$

代入 (1.5) 得

$$\sqrt{n}k_1k_2\cdots k_{n-1} \geq \frac{\sqrt{n!}}{\sqrt{(n-l(n))!}}\exp\left(-24\sqrt{n\log n}\right)\prod_{j=l(n)+1}^{n-1}k_j. \tag{1.6}$$

另一方面, 对所有 $i > l(n)$ 取 $k_i = \dfrac{1}{n^2}$, $l(n) = n - \log\log n$, 使得 $n - l(n) \to +\infty$, 即目标 2 成立. 同时, 当 n 充分大时, 有

$$\frac{\prod_{j=l(n)+1}^{n-1}k_j}{\sqrt{(n-l(n))!}} \geq \frac{n^{-2(n-l(n))}}{\sqrt{(n-l(n))!}} \geq \exp\left(-\sqrt{n\log n}\right).$$

由此代入 (1.6) 可知

$$\sqrt{n}k_1k_2\cdots k_{n-1} \geq g(n)\sqrt{n!}.$$

这里 $g(n) = \exp\left(-29\sqrt{n\log n}\right)$ 已优化至 $\exp\left(-25\sqrt{n\log n}\right)$. 现在我们有如下命题:

命题 1. 当 n 充分大时, 对 $\forall i$ 满足 $l(n)+1 \leq i \leq n-1$, 我们有

$$\mathbb{P}\left(d(X_{i+1}, W_i) \leq \frac{1}{n^2}\right) = O\left(\frac{1}{\sqrt{\log n}}\right).$$

由该命题可知,

$$\sum_{i=l(n)+1}^{n}\mathbb{P}(d(X_{i+1}, W_i) < k_i) = O\left(\frac{1}{\sqrt{\log n}}\right)(n-l(n)) = O\left(\frac{1}{\sqrt{\log n}}\right)(\log\log n) \xrightarrow{n\to+\infty} 0.$$

结合 (1.4) 可知 (1.3) 成立, 从而 $\lim\limits_{n\to+\infty}\mathbb{P}\left(|\det M_n| \geq g(n)\sqrt{n!}\right) = 1$. 现在我们只需完成上述命题的证明.

命题 1 证明: 仅考虑 $i = n-1$ 的情形, 其余类似. 这里我们证明对任意固定的 X_n, 待证命题均成立. 对 $\boldsymbol{w} = (w_1, \cdots w_n)$, 记样本空间 $\Omega = \left\{\boldsymbol{w} \in W_{n-1}^{\perp}, |\boldsymbol{w}| = 1 \text{且} X_n \cdot \boldsymbol{w} = d(X_n, W_{n-1})\right\}$(后续均在 $\boldsymbol{w} \in \Omega$ 的意义下考虑), 仅需证明: 当 n 充分大时, 有

$$\mathbb{P}\left(\left\{\boldsymbol{w} \in \Omega : |e_{n1}w_1 + \cdots + e_{nn}w_n| \leq \frac{1}{n^2}\right\}\right) = O\left(\frac{1}{\sqrt{\log n}}\right). \tag{1.7}$$

定义事件

$$A_t = \left\{\boldsymbol{w} \in \Omega : \exists 1 \leq i_1 < \cdots < i_t \leq n, \text{s. t. } |w_j| \geq \frac{2}{n^2}, \forall j \in \{i_1, \cdots i_t\}\right\},$$

一方面,

$$\mathbb{P}\left(A_t \bigcap \left\{\boldsymbol{w} \in \Omega : |e_{n1}w_1 + \cdots + e_{nn}w_n| \leq \frac{1}{n^2}\right\}\right) \tag{1.8}$$
$$\leq \sup_{x\in\mathbb{R}}\mathbb{P}\left(\left\{\boldsymbol{w} \in \Omega : |e_{ni_1}w_{i_1} + \cdots + e_{ni_t}w_{i_t} - x| \leq \frac{1}{n^2}\right\}\right)$$
$$= \sup_{x\in\mathbb{R}}\mathbb{P}\left(\left\{\boldsymbol{w} \in \Omega : \frac{n^2}{2}(e_{ni_1}w_{i_1} + \cdots + e_{ni_t}w_{i_t}) \in [x-1, x]\right\}\right). \tag{1.9}$$

我们要利用组合学课程中学过的 Littlewood-Offord Problem, 即:

引理 2 (Littlewood-Offord Problem). 固定向量 $\boldsymbol{a} = (a_1, \cdots, a_n)$, 且对任意 $1 \le i \le n$, $|a_i| \ge 1$. 令

$$S = \{\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots \varepsilon_n) : \varepsilon_i \in \{-1, 1\}, \boldsymbol{\varepsilon} \cdot \boldsymbol{a} \in I\},$$

其中 I 是长度小于 2 的区间, 则

$$|S| \le \binom{n}{[\frac{n}{2}]}.$$

因为

$$\left|\frac{n^2}{2} w_{i_j}\right| \ge 1, \quad 1 \le j \le t.$$

利用 Littlewood-Offord Problem, 结合 Stirling 公式: $n! \sim \sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n$, 代入 (1.9) 可得

$$\mathbb{P}\left(A_t \bigcap \left\{\boldsymbol{w} \in \Omega : |e_{n1}w_1 + \cdots + e_{nn}w_n| \le \frac{1}{n^2}\right\}\right) \le \frac{1}{2^t}\binom{t}{[\frac{t}{2}]} = O\left(\frac{1}{\sqrt{t}}\right).$$

另一方面, 我们定义事件 $B_t = \left\{\boldsymbol{w} \in \Omega : w_i \ge 0, 1 \le i \le t-1; |w_j| < \frac{2}{n^2}, t \le j \le n\right\}$, 则有

$$\mathbb{P}(A_t^c) \le \binom{n}{t-1} \cdot 2^t \cdot \mathbb{P}(B_t). \tag{1.10}$$

当 $\boldsymbol{w} \in B_t$ 时, 注意到对任意 $1 \le s \le n-1$, 有

$$e_{1s}w_1 + e_{2s}w_2 + \cdots + e_{ns}w_n = 0.$$

因此

$$|e_{1s}w_1 + \cdots + e_{t-1,s}w_t| = |e_{ts}w_t + \cdots + e_{ns}w_n| \le |w_t| + \cdots + |w_n| \le \frac{2(n-t+1)}{n^2}.$$

而

$$\begin{aligned} w_1 + \cdots + w_{t-1} &\ge |w_1|^2 + \cdots + |w_{t-1}|^2 = 1 - (|w_t|^2 + \cdots + |w_n|^2) \\ &\ge 1 - \frac{4(n-t+1)}{n^4} > \frac{2(n-t+1)}{n^2}, \end{aligned}$$

故对任意 $1 \le s \le n-1$, $\left\{e_{sj}\right\}_{j=1}^n$ 不全为正数, 由此知 $\mathbb{P}(B_t) \le (1-2^{1-t})^{n-1}$, 代入 (1.10) 得

$$\mathbb{P}(A_t^c) \le \binom{n}{t-1} \cdot 2^t \cdot (1-2^{1-t})^{n-1} \le (2n)^t(1-2^{1-t})^{n-1} \le (2n)^t \mathrm{e}^{-2^{1-t}(n-1)}.$$

取 $t = \frac{\log n}{2}$, 则有

$$\begin{aligned} &\mathbb{P}\left(\left\{\boldsymbol{w} \in \Omega : |e_{n1}w_1 + \cdots + e_{nn}w_n| \le \frac{1}{n^2}\right\}\right) \\ \le\ &\mathbb{P}\left(A_t \bigcap \left\{\boldsymbol{w} \in \Omega : |e_{n1}w_1 + \cdots + e_{nn}w_n| \le \frac{1}{n^2}\right\}\right) + \mathbb{P}(A_t^c) \\ \le\ &O\left(\frac{1}{\sqrt{t}}\right) + (2n)^t \mathrm{e}^{-2^{1-t}(n-1)} = O\left(\frac{1}{\sqrt{\log n}}\right), \end{aligned}$$

其中最后一步是因为

$$\log\left((2n)^t \mathrm{e}^{-2^{1-t}(n-1)}\right) = \frac{1}{2}\log n \log(2n) - 2^{1-\frac{1}{2}\log n}(n-1) \le \frac{1}{2}\log n \log(2n) - \frac{2(n-1)}{\sqrt{n}}.$$

因此 (1.7) 成立, 命题得证.

(5)a.s. 下成立并未解决. 下仅证明依概率收敛. 由 (3)(4) 知

$$\lim_{n\to+\infty}\mathbb{P}\left(\exp\left(-29\sqrt{n\log n}\right)\sqrt{n!}\le|\det M_n|\le\exp\left(\sqrt{n\log n}\right)\sqrt{n!}\right)=1.$$

结合 Stirling 公式: $n!\sim\sqrt{2\pi n}\left(\dfrac{n}{\mathrm{e}}\right)^n$, 可知

$$\frac{\log|\det M_n|}{n\log n}\xrightarrow{P}\frac{1}{2}.$$

编者按: 此解答来自宗语轩同学.

致谢

在科大数院、少院的各位老师、校友、同学们，其他数学爱好者、数学工作者及校内外热心人士的共同努力与帮助下，《蛙鸣》复刊的第三期——第 66 期《蛙鸣》得以成功发行。在此，编辑部全体成员非常感谢各位投稿人的踊跃来稿，感谢第九届数院学生会的同学们参与排版设计与宣传发放工作，感谢科大制本厂对本期《蛙鸣》的印刷与制作工作的全力支持。

编辑部在此感谢参与本期审稿的各位审稿人（顺序不分先后）：刘群欢、李书亮、聂思捷、赵理浩、杨威、何家志、宗语轩、杜哲航、鲍越恒、周可树、卢昊翔、吴佳骏、荆楷豪、陈陌青、叶星辰、孙孝同。感谢他们在学习和科研的百忙之中抽出时间完成了谨慎、专业的审稿工作。感谢章俊彦校友、陈恒宇校友对本期《蛙鸣》的指导。感谢林开亮老师一直以来对《蛙鸣》的热情关怀与鼓励。

编辑部在此感谢中国科学技术大学数学科学学院对第 66 期《蛙鸣》给予的全方位支持。编辑部还特别致谢马杰老师、王作勤老师、郑芳老师对本期《蛙鸣》的筹备与编撰工作的悉心关怀与指导。

最后，感谢各位读者阅读本期《蛙鸣》，期待于明年夏天与各位再见！

第 66 期《蛙鸣》编辑部

2023 年 7 月 20 日

《蛙鸣》第 67 期征稿启事

现在，我们正式为《蛙鸣》第 67 期征稿！

创刊宗旨

《蛙鸣》是中国科大数院的学生杂志。1981 年 6 月 20 日，首期《蛙鸣》由 78 级数学系的同学们自写、自编、自刻、自印而成。四十余年来，《蛙鸣》一直是一个完全由学生主导，共同探讨、自由交流数学的开放平台，让同学们可以互相交流彼此的思想和发现。所以，我们欢迎各位科大的校友和同学们踊跃投稿！同时，我们也欢迎外校师生投稿，增进交流！

投稿栏目

(1) **蛙鸣记忆** 记录属于你的科大故事，叙述你与科大和数学的点点滴滴。

(2) **初阳** 以小论文形式研究一个无需太高深的数学知识就足够理解和欣赏的问题，或对具有一定深度的数学内容进行深入浅出的科普。欢迎大一大二同学来稿！

(3) **星辰** 分享你所喜爱或崇敬的数学工作者的故事、演讲与采访等，或让你印象深刻的讲座、座谈会与研讨会等的记录，或由你所撰写、翻译的数学科普与数学历史等。

(4) **蛙声一片** 以小论文形式记录你对数学更深入的学习或研究，可以撰写关于一个前沿方向或经典问题与相关成果的综述，可以研究一个具有一定深度与背景的数学问题，也可以记录你对数学课程的进阶内容的学习与思考。

(5) **笔墨诗篇** 分享你的原创文学作品，体裁不限。

投稿步骤

请将稿件发送到数院学生会官方邮箱：`mathsu01@ustc.edu.cn`

(1) 邮件标题：蛙鸣 _ 投稿栏目 _ 文章标题 _ 作者。

(2) 邮件正文：作者信息、联系方式（邮箱、QQ 或微信等）。

(3) 邮件附件：稿件 PDF（请隐去作者信息，以便盲审）。确定录用后请发送稿件 tex 源码。

文章要求

(1) 含有较多公式的文章请使用 LaTeX 编写，推荐使用《蛙鸣》官网模板。

(2) 数学类文章可以使用中文或英文撰写，要求语言朴实流畅，无明显语病。

(3) 文章摘要：请在正文前简要介绍本篇文章用到的想法、思路、主要内容等概要信息。

(4) 非原创内容请注明参考文献（尽可能详细到章节）。

(5) 其他投稿相关问题可以加入第 67 期蛙鸣咨询群（892896803）咨询。

更多资讯请前往《蛙鸣》官网 `http://staff.ustc.edu.cn/~mathsu01/pu/waming.html` 查看，或关注数院学生会官方 QQ（2061453364）与微信公众号（中国科大数学科学学院学生会）了解。

感谢大家对蛙鸣的支持！

下期预告

截至目前，《蛙鸣》第 67 期已征得稿件有：

Asymptotic Expansion and Estimates of Bergman Kernel 田珺昊

欢迎大家踊跃来稿！

《蛙鸣》主页

数院学生会 QQ 号

数院学生会微信公众号

其形虽小
其声也宏
充实基础
奏出强音

谷超豪